



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Ultra-fast Coherent Diffraction Imaging of Single Particles, Clusters and Biomolecules

J. Hajdu, H. Chapman

October 7, 2006

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

ULTRA-FAST COHERENT DIFFRACTION IMAGING OF SINGLE PARTICLES, CLUSTERS, AND BIOMOLECULES

Janos Hajdu (1,2) and Henry N. Chapman (3)

(1) Laboratory of Molecular Biophysics, Institute of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, S-75124 Uppsala, Sweden

(2) Stanford Linear Accelerator Centre, Stanford University, MS69, 2575 Sand Hill Road, Menlo Park, California 94305, USA

(3) University of California, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore CA 94550, USA

SUMMARY

Theoretical studies and simulations predict that with a very short and very intense coherent X-ray pulse a single diffraction pattern may be recorded from a large macromolecule, a virus, or a cell without the need for crystalline periodicity [Neutze *et al.*, 2000; Jurek *et al.*, 2004; Hau-Riege, 2004 *et al.*; Bergh *et al.*, 2004]. Measurement of the over-sampled X-ray diffraction pattern permits phase retrieval and hence structure determination [Bates, 1982; Fienup, 1982; Sayre *et al.*, 1998; Miao *et al.*, 1999; Robinson *et al.*, 2001; Marchesini *et al.*, 2003; Chapman, 2006]. Although individual samples will be destroyed by the very intense X-ray pulse, a three-dimensional data set could be assembled when copies of a reproducible sample are exposed to the beam one by one [Huldt *et al.*, 2003]. The challenges in carrying out such an experiment requires an interdisciplinary approach, drawing upon structural biology, atomic and plasma physics, mathematics, statistics, and XFEL physics. The potential for breakthrough science in this area is great with impact not only in the biological areas but wherever structural information at or near atomic resolution on the nanoscale is valuable.

1. THE SCIENTIFIC CASE FOR ULTRA-FAST DIFFRACTION IMAGING

Diffraction imaging is elegant in its experimental simplicity: a coherent X-ray beam illuminates the sample and the far-field diffraction pattern of the object is recorded on an area detector. The scientific case for the ultra-fast imaging of single particles, biomolecules, and clusters is based on published theoretical studies and simulations [Neutze *et al.*, 2000; Hajdu & Weckert, 2001; Hajdu, 2002; Jurek *et al.*, 2004; Hau-Riege, 2004 *et al.*; Bergh *et al.*, 2004], and is supported by recent experimental results (discussed below) from the first soft X-ray free-electron laser, the *FLASH* facility at DESY (formerly called the VUV-FEL).

Radiation damage significantly limits the resolution of conventional imaging experiments. Damage is caused by energy deposited into the sample by the probes used for imaging (photons, electrons, neutrons, etc.). Cooling can slow down sample deterioration, but it cannot eliminate damage-induced sample movement *during* conventional measurements [Henderson, 1990; 1995; Nave 1995]. Ultra short x-ray pulses from X-ray free-electron lasers offer the possibility to extend the conventional damage limits, and will allow the imaging of non-crystalline biological (and other) materials. For proteins, simulations based on molecular dynamics (MD) [Neutze *et al.*, 2000; Jurek *et al.*, 2004a, 2004b], hydrodynamic [Hau-Riege *et al.*, 2004], and on plasma models [Bergh *et al.*, 2004] indicate that if very short (100 fs or less) and very intense x-ray pulses are available ($\geq 10^6$ photons/Å² on the sample), then a single scattering pattern could be recorded from a single protein molecule in the gas phase before radiation damage manifests itself and ultimately destroys the sample (Figure 1).

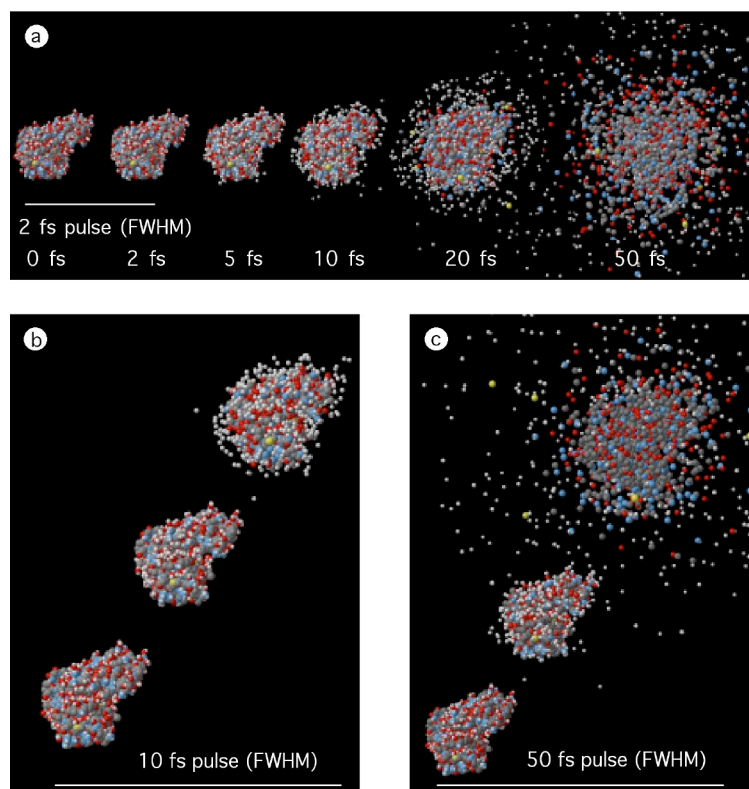


Figure 1: MD simulation of radiation-induced Coulomb explosion of a small protein (lysozyme). White balls: H, Gray: C, Blue: N, Red: O, Yellow: S. Integrated X-ray intensity: 3×10^{12} (12 keV) photons/100 nm diameter spot (corresponding to 3.8×10^8 photons/nm², or 3.8×10^6 photons/Å² on the sample) in all cases. (a) Protein exposed to a 2 fs FWHM X-ray pulse, and disintegration followed in time. The atomic positions in the first two structures (before and after the pulse) are practically identical at this pulse length due to an inertial delay in the explosion. (b) Lysozyme exposed to the same number of photons as in (a) but the pulse FWHM is now 10 fs. The images show the structure at the beginning, in the middle and near the end of the X-ray pulse. (c) Behaviour of the protein during a 50 fs FWHM X-ray pulse. It is also apparent from the figure that during the Coulomb explosion, hydrogen ions and highly ionised sulphurs are the first to escape the immediate vicinity of the protein (at 12 keV, the photoelectric cross section for sulphur is about fifty times larger than that for carbon). Based on Neutze et al. [2000].

Only a small fraction of the x-rays that fall on the sample will actually interact with the molecule. The “useful” photons are those that scatter elastically and produce the diffraction pattern. However, the dominant interaction is K-shell photoionisation of the atoms in the molecule, which leads to damage. It produces electrons with energy close to the incident photon energy (around 12 keV in the examples here). Several femtoseconds later, the K-shell holes of the non-hydrogenic atoms (C, N, O, etc.) decay, mainly by emitting Auger electrons with relatively low energies 250 – 500 eV. In biologically relevant elements (C, N, O, S, P), a single absorbed X-ray photon leads to more than a single ionisation event this way. The photoelectrons and the Auger electrons have different energies, and are released at different times. Early on in an exposure, most of these electrons escape the molecule. The escaping electrons leave behind a net positive charge. This charge can grow large enough during an exposure so that the less energetic Auger electrons become electrostatically trapped in the molecule. The more energetic photoelectrons become trapped later and only in larger molecules. Both escaping and trapped photoelectrons produce secondary electrons through collisional ionization of atoms. The escaping photoelectrons produce at most a few secondaries, while trapped Auger and photoelectrons produce the majority of the secondaries. This has been modeled in detail for large (and cold) carbon samples by Ziaja et al. [2002; 2003; 2005; 2006], and for water clusters by Timneanu et al. [2004]. Figure 1 illustrates conditions near the physical limits in imaging [Neutze et al., 2000]. For the two shorter X-ray pulses only very small changes in the atomic positions have had time to develop during the exposure (Figures 1a and b) although most of the atoms in the sample became photo-ionised. The life-times of K-shell core holes in biologically

relevant elements (C, O, N) is long (around 10 fs), and as a consequence, very little Auger emission takes place in the sample during exposures that are shorter than about 10 fs (see below).

Resolution in these diffraction experiments does not depend on sample quality in the same way as in conventional crystallography, but is a function of radiation intensity, pulse duration, wavelength, and the extent of ionization and sample movement during the exposure [Neutze et al., 2000; Jurek et al., 2004a, 2004b; Hau-Riege et al., 2004; Bergh et al., 2004].

There is an additional component that will influence the final resolution of a three-dimensional diffraction data set. Each particle (macromolecule) is exposed to the beam only once, and disintegrates at the end of this process. The diffraction pattern so recorded encodes a two-dimensional projection image of the sample (and this may provide sufficient information for some applications). Three-dimensional imaging requires more than one view from the sample. In addition, the signal-to-noise ratio of raw diffraction images will probably be insufficient for a high-resolution reconstruction, and it will be necessary to obtain a redundant data set so that averaging can enhance the signal. One could extend the depth of view from a single exposure by various holographic techniques based on external or internal reference beams, but a full three-dimensional reconstruction will most likely require *reproducible samples* exposed to the beam one-by one, and in different orientations. A “reproducible sample” (e.g. purified proteins) may contain heterogeneities, different subgroups of sample, and distinct conformers of the molecule. How reproducible is a “reproducible sample” and how well can we distinguish between similar and dissimilar structures will affect resolution through a B-factor-like component.

Conventional “single molecule” electron cryo-microscopy [Frank, 1996; van Heel et al. 2000] faces similar challenges as those described here. The basic requirement for reconstruction and/or signal averaging from many diffraction images is the ability to tell whether two noisy diffraction patterns represent the same view of the sample or two different views [Huldt et al. 2003]. With this knowledge, averaging techniques can be used to enhance the signal and extend the resolution in a redundant data set. A 3D diffraction reconstruction from a number of different exposures and views can take various routes. Techniques from “single molecule” electron cryo-microscopy [Frank, 1996; van Heel et al. 2000] can be adapted here for the relative orientation and merging of a large number of noisy two-dimensional diffraction patterns.

In the next stage, phases are derived from the 3D diffraction pattern. In principle, there is a direct way to determine the phases required for the conversion using an approach called “oversampling” [for a recent review see Sayre, 2002], and this has now been used successfully in a number of experiments [see, e.g. Miao et al., 2001; 2002; 2003; Robinson et al., 2001; Marchesini et al., 2003a,b]. Practical algorithms exist [Marchesini et al., 2003a] that are very robust and allow *ab initio* reconstruction of 3D structures.

In practical terms, there is a need for “containerless packaging” so that only the sample of interest is imaged. We will also need sensitive detection of diffraction patterns with noise levels less than a single photon per pixel (discussed below), and the development of advanced reconstruction algorithms for orienting [Huldt et al. 2003], averaging, and inverting the diffraction images into the real space molecule [e.g. Marchesini et al., 2003a,b].

1.1 Photon-material interactions on ultra-short time scales and at high x-ray intensities

Energy absorption is a subject of interest in practically all applications of XFELs, from biological imaging to creating astrophysical conditions in the laboratory. Imaging of single particles and biomolecules is based on the use of extremely intense and extremely short X-ray pulses to limit

radiation induced changes *during* the exposure so that high quality structural data is collected before significant changes take place in the sample (Figure 1). Single molecule imaging experiments at XFELs represents the high-end of today's high-energy density science. In a focused beam of an XFEL, more than 100 eV/atom will be deposited into the sample within a few femtoseconds, and this will turn the sample into a plasma at some point. The question is when and how?

Figure 2 shows expected behaviour of an atom in a very intense photon field at optical and at X-ray frequencies.

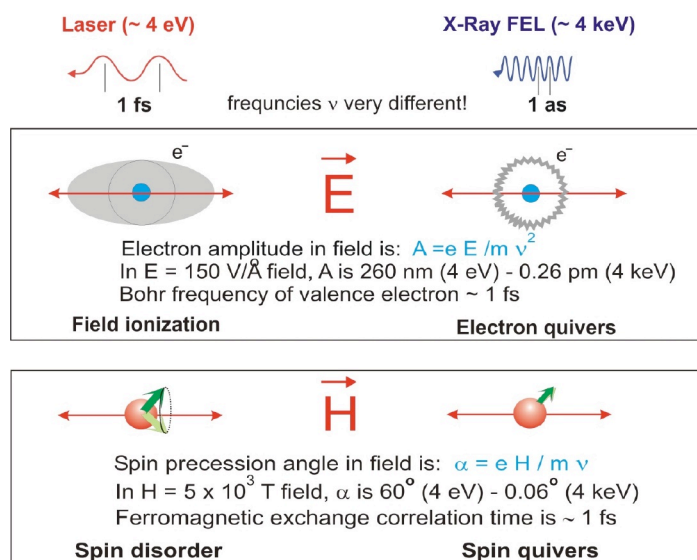


Figure 2. Electrons and spins in high photon fields.

Under extreme conditions in an intense photon field, electrons can be stripped rapidly from an atom. At optical frequencies, the high-field regime begins to occur at intensities of about $10^{14-15} \text{ W/cm}^2$, where some of the new atomic physics and plasma physics phenomena have been observed. At around 10^{18} W/cm^2 , relativistic effects fully enter the dynamics of electrons in the optical frequency domain. The high-field regime in optical frequencies is well known but it is completely unexplored at X-ray frequencies (no experiments have so far been possible due to a lack of suitably intense X-ray sources). X-rays have 3-4 orders of magnitude higher frequencies than optical light, and as a consequence, the high-field limits for X-rays are expected to be much higher than those for the slower optical frequencies. Various estimates suggest that field ionization becomes significant at around $10^{22-23} \text{ W/cm}^2$, and relativistic effects will dominate the picture at around 10^{26} W/cm^2 . These numbers are very high compared to expected intensities of unfocused pulses from an XFEL ($10^{16-17} \text{ W/cm}^2$). However a highly focused X-ray beam ($<100 \text{ nm}$ focal spot) with more than 10^{12} photons in a pulse, and at pulse durations shorter than about 5 fs, will approach the high field limits for X-rays. A quantum mechanical analysis of the electric field induced tunneling indicates that atoms may become stabilised against ionisation under these conditions in a high-frequency X-ray field [Reiss, 1997]; however, there are no experimental data near this regime. Expected stabilisation effects and other non-linear effects are not included in current damage models.

It is important to point out that other major components of the mechanism of damage formation are also different at optical and X-ray frequencies. At optical frequencies, outer shell processes dominate the ionisation of a sample; while at X-ray frequencies inner shell processes take over (photoemission followed by Auger emission/X-ray fluorescence, shake-up and shake-off excitations, etc.). Inverse Bremsstrahlung absorption is significant at optical frequencies, while it is small at X-ray frequencies. These properties affect the evolution of Coulomb explosions, and are responsible for differences in the heating of the sample during (and after) a short exposure.

1.1.1 Physics of damage formation with X-rays

At 12 keV X-ray energy (~ 1 Å wavelength), the photoelectric cross section of carbon is about 10 times higher than its elastic scattering cross section making the **photoelectric effect** the primary source of damage. During this process a photon is absorbed and an electron is ejected usually from a low-lying orbital of the atom (about 95% of the photoelectric events remove K-shell electrons from carbon, nitrogen, oxygen and sulphur), producing a hollow ion with an unstable electronic configuration. Relaxation is achieved through a higher shell electron falling into the vacant orbital. In heavy elements, this usually gives rise to X-ray fluorescence, while in light elements, the electron falling into the lower orbital is more likely to give up its energy to another electron, which is then ejected in the **Auger process**. Auger emission is predominant in light elements like carbon, nitrogen, oxygen and sulphur (99-95%), thus, most photoelectric events ultimately remove two electrons from these elements. These two electrons have *different energies*, and are released at *different times*. Relevant K-hole life-times can be determined from Auger line-widths, and are 11.1 fs (C), 9.3 fs (N), 6.6 fs (O) and 1.3 fs (S). Note that the chemical environment of an atom will influence Auger life times to some degree from isolated atom values. **Shake-up and shake-off excitations** (multiple ionisation following inner shell ionisation, see, e.g. Persson et al., 2001), initial- and final-state configuration interaction and interference between different decay channels will modulate this picture. The chemical environment of atoms influences shake-effects to some degree. The release of the unbound electron "competes" with Auger electrons. When the first electron velocity is low (at low X-ray energies), the slow electron can interact with the other (valence) electrons on its way out and exchange energy. If the energy of the first electron is above some threshold, the sudden approximation is valid, and in general terms, less shakeup will happen. We expect this to be the case with primary photoelectrons of about 12 keV energies, and only a small shakeup fraction is expected (about 10%, Persson et al., 2001). **Inelastic scattering** represents a direct momentum transfer from an X-ray photon to an electron, so that the X-ray photon is scattered with an altered energy. If the energy taken up by a bound electron is greater than its shell binding energy, the atom will be ionised. The inelastic cross section of carbon, nitrogen and oxygen is around 3% of the corresponding photoelectric cross sections, whereas the inelastic cross section of hydrogen is much higher than its photoelectric cross section. Electrons ejected from atoms during exposure propagate through the sample, and cause further ionisation by eliciting **secondary electron cascades**. The extent of ionisation through this mechanism will depend on the size of the sample. Photoelectrons released by X-rays of 1 Å wavelength are fast (660 Å/fs), and they can escape from small samples early in an exposure. In contrast, Auger electrons are slow (95 Å/fs) for carbon and it is likely that they will thermalise even in a small sample. A detailed description of secondary electron cascades initiated by an electron with impact energies between 0.1 and 10 keV has been published [Ziaja et al., 2001; 2002; 2005; 2006]. In *late phases* of an exposure, a significant fraction of the emitted electrons will not be able to escape the increased positive potential of the sample. **Trapped electrons** will increase the kinetic energy of the sample through **thermal equilibration**, while at the same time they will also slow down the Coulomb explosion of the sample by **partially screening** the positively charged protein core (all plasmas start cold and dense and become hot later). These opposing effects are explicitly incorporated into the damage models of Bergh et al. [2004], Hau-Riege et al. [2004], and Jurek et al. [2004], and are balanced by simplifying assumptions in the model of Neutze et al. [2000]. Finally, a **transient radiation hardening of the sample** can be expected **during very short exposures**. The dominant interaction of hard X-rays with atoms is through K-shell photoionisation. This process creates hollow ions with one or both of the K-electrons expelled from the atom. The probability of photoionisation by X-rays in hollow ions is expected to be smaller than in relaxed ions or intact atoms. The short-lived "ionisation-resistance" lasts until outer shell electrons fill the vacant K-hole(s) (an exponential decay with about 10 fs half life in C, N, O), and one may expect a gradually increasing stability for biomolecules in the beam during short exposures (≤ 10 -20 fs).

1.1.2 Modelling damage formation and sample dynamics

A number of damage models have been developed during the past few years [Neutze et al., 2000; Persson et al., 2000; Ziaja et al., 2001; 2002; 2005; 2006; Timneanu et al., 2004; Bergh et al., 2004; Jurek et al., 2004a, 2004b, Hau-Riege et al., 2004]. Computer simulations using four different models [Neutze et al., 2000; Jurek et al., 2004a, 2004b; Hau-Riege et al., 2004; Bergh et al., 2004] are in general agreement with each other, and suggest that the structure of a molecule could be determined by judicious choice of FEL pulse length, intensity and wavelength before it is stripped by electrons, and is destroyed in a Coulomb explosion.

The molecular dynamics (MD) model of Neutze et al. [2000] describes X-ray induced damage stochastically based on the probability of a photoelectric or an inelastic event. The force field incorporates Morse potentials for the description of chemical bonds, thereby enabling bonds with sufficiently high energy to break. For water, a simple point charge model was used and adapted in the same manner. The instantaneous probability of ionisation of atom j at time t is calculated as the product of its photoelectric cross section, and the time-dependent intensity of the X-ray pulse, $I(t)$. Auger emission is modeled as a stochastic exponential decay to reproduce appropriate K-hole lifetimes. The direction of photo-emission is distributed according to a random deviate which follows a Gaussian distribution. A recoil velocity for the ionised atom due to inelastic scattering or the emission of a photo- or Auger electron is calculated from energy and momentum conservation. For inelastically scattered photons, the angle of deflection is determined by a random deviate following a Rayleigh distribution. For each inelastic scattering event, the electron's recoil energy is calculated, and in cases when this is greater than the binding energy of the electron, ionisation is modeled (this approximation was deemed sufficient for light elements with weakly bound electrons). An inventory is kept on all electrons in the sample, and changes in the photoelectric, elastic and inelastic scattering cross sections are computed for all atoms, hollow ions and relaxed ions in 50 attosecond periods during exposure. Trapped electrons will increase the kinetic energy of the sample through thermal equilibration, but they will also slow down the Coulomb explosion of the sample by partially screening the positively charged protein core (all plasmas start cold and dense and become hot later). These opposing effects are balanced by simplifying assumptions in the model of Neutze et al. [2000].

The hydrodynamic model (HD) of Hau-Riege et al. [2004a] includes the trapping of photoelectrons but does not treat atomicity explicitly. This model can be applied to both very small and very large samples (millimetres) while MD models are limited to macromolecules because of computing costs. The basic assumption of the HD model is that the sample can be described by a liquid-like continuum of matter rather than considering individual atoms. This gives a simplified description of the average effects of x-ray material interaction and atomic motion, which then permits calculations even on very big samples. The model further assumes that the particle is spherically symmetric, reducing the mathematical model to *one dimension plus time*. The model assumes that the motion of the atoms within the molecule is solely in the radial direction. The electrons and the atoms are treated as separate, structureless, fluids that interact through the Coulomb force and ionization processes. The short-range electron-electron interactions are treated as a hydrodynamic pressure, and the long-range electron-electron and electron-ion Coulomb interactions are determined from the continuous net charge of the electrons and ions. In this model, all forces act radially. The model further assumes that the trapped electrons are thermalised among themselves, and that they are inertia free, so that they quickly relax to a force-free spatial equilibrium. Finally, the x-ray matter interaction, atomic ionization processes, and energy of the trapped electrons are described by time-dependent rate equations. The model shows that at later phases in an exposure, trapped electrons quickly relax in energy and position to form a cloud around the positive ions, leaving a neutral core and a positively charged outer shell (similar to Debye shielding). It is this positively charged outer shell that peels off first, and the Coulomb explosion then “burns” from the outside towards the inside (similar behaviour can be observed in the other models). In the inner core, there is hardly any ion motion but the high electron temperature leads to ionization and blurring of the electron density. It is this latter effect that requires short pulse lengths to overcome damage.

Two further models have been published recently to include electron-electron, electron-atom, and electron-ion interactions:

Molecular dynamics with explicit electrons [Jurek et al. 2004] - a molecular dynamics/Monte Carlo model (MDMC). This model extends the GROMACS-based MD model of Neutze et al. [2000] by including explicit electrons in the simulations. Numerical modeling is based on the non-relativistic classical equation of motion. Quantum processes are taken into account by the respective cross-sections. Just like the first MD model, this model also includes photon-electron, photon-atom, and photon-ion cross-sections in addition to a description of the atom-atom, atom-ion, ion-ion, interactions. The MDMC model extends this by treating explicitly electron-atom, electron-ion, and electron-electron interactions. The model gives very detailed information on the movement of all atoms and electrons in and around the sample. A disadvantage of the model is that it is extremely demanding on computing power, and this limits its use to small samples only (a couple of thousand atoms).

Molecular dynamics with implicit electrons [Bergh *et al.*, 2004] - a molecular dynamics/continuum electron model (MDCE). This model uses a plasma approximation for the description of electrons released and trapped in the sample, and extends the GROMACS-based model of Neutze et al. [2000] to include the effect of screening by free electrons through the inclusion of an electron gas. The electrons are approximated by a classical gas, and the electron distribution is calculated iteratively from the Poisson-Boltzmann equation. Simulations of water clusters reveal the details of the explosion dynamics, as well as the evolution of the free electron gas during the beam exposure. Inclusion of the electron gas in the model slows down the Coulomb explosion. Hydrogen atoms leave the sample faster than oxygen atoms, leading to a double layer of positive ions. A considerable electron density is located between these two layers. The fact that protons leave much faster than the oxygens means that the heavy part of the sample stays intact somewhat longer than the sample as a whole. A disadvantage of this model is its demand on computing power due to expensive grid calculations when simulating larger samples.

While these models are significantly different from each other, they all come up with similar predictions.

1.1.3 Calculated scattering patterns for single molecules and particles

Radiation damage causes changes to atomic scattering factors (mainly through ionisations) and also to atomic positions. The creation of a large number of positive charges in close proximity within the sample results in a rise in the electrostatic energy of the sample, which drives its eventual explosion (Figure 1). The degree of conversion of potential energy into kinetic energy during the X-ray exposure is inertia limited, and as a consequence, strongly depends on the duration of the pulse and on the forces that develop. During the 2 fs pulse in Figure 1a, there was insufficient time for the kinetic energy to grow appreciably. In contrast, by the completion of the 50 fs pulse (Figure 1c), the kinetic energy of the sample had surpassed its potential energy, and the explosion of the sample was well under way already during the exposure.

The effect of ionisation and ionisation-induced sample movement on the diffraction pattern can be calculated exactly at any point along the simulated trajectories. For unpolarised X-rays, the mean number of elastically scattered photons $I(\mathbf{u}, \Omega)$ from a molecule illuminated by a plane wave, to be detected by an idealised detector pixel of a projected solid angle Ω centred at a positional vector \mathbf{u} , is given by

$$I(\mathbf{u}, \Omega) = 1/2(1 + \cos^2 2\theta) \Omega r_e^2 \int_{-\infty}^{\infty} I(t) \left| \sum_j f_j(t) \exp\{i\Delta\mathbf{k}(\mathbf{u}) \cdot \mathbf{x}_j(t)\} \right|^2 dt$$

where r_e is the classical electron radius; $I(t)$ is the intensity of the X-ray pulse; $f_j(t)$ is the atomic

scattering factor for the j^{th} atom as a function of time; $\mathbf{x}_j(t)$ is the position of this atom as a function of time; and $\Delta\mathbf{k}$ is the change in the wave vector of the X-ray photon when scattered through 2θ radians towards the pixel centred at \mathbf{u} .

Radiation damage interferes with the atomic scattering factors $f_j(t)$ and the atomic positions $\mathbf{x}_j(t)$. Damage-induced changes in the scattering pattern, and the structural information, which can be recovered from the sample during an exposure can thus be quantified. Assessment of damage-induced changes in the diffraction pattern is based on the calculated differences between scattering from a sample that suffers radiation damage (I_{real}), and scattering from a hypothetical sample that suffers no radiation damage (I_{ideal}) over the same time period. The damage-induced error in the integrated diffraction pattern can thus be described by a weighted average agreement factor (R) between the two patterns as defined below:

$$R = \sum_u \left| \frac{K^{-1} \sqrt{I_{\text{real}}(u)} - \sqrt{I_{\text{ideal}}(u)}}{\sum_{u'} \sqrt{I_{\text{ideal}}(u')}} \right| \quad K = \frac{\sum_u \sqrt{I_{\text{real}}(u)}}{\sum_u \sqrt{I_{\text{ideal}}(u)}}$$

$I_{\text{real}}(u)$ is derived from the time-dependent atomic coordinates, $\mathbf{x}_j(t)$, and scattering factors, $f_j(t)$, of a sample exploding in the X-ray pulse, while $I_{\text{ideal}}(u)$ is determined from a reference molecular dynamics simulation of an unexposed sample. R provides information on the extent to which the elastically scattered radiation is perturbed by X-ray-induced damage and provides a direct assessment of data quality ($R = 0$ is ideal, larger R means larger errors). Scaling factor K describes the relative scattering power of the sample. Macromolecular crystal structures in the Protein Data Bank have crystallographic R factors of about 20%. Many of the structures, especially those collected earlier on photographic film, represent data sets with merging R factors in the 5% to 15% range. Taking the latter value as an arbitrary upper limit, we regard damage as acceptable if $R \leq 15\%$.

1.1.4 The landscape of damage tolerance

The four published damage models are fairly different in what they include and what they neglect in their description of atoms, ions, electrons and their interactions with an intense X-ray pulse, but all models arrive at surprisingly similar results about pulse, fluence, and wavelength requirements for diffraction imaging. Here we pick one of these models [Neutze et al., 2000] to illustrate the available phase space for high-resolution imaging of single macromolecules.

Figures 3 and 4 show the landscape of damage tolerance in a broad parameter space around the expected pulse parameters of XFELs. Figure 3 gives calculated R -values for a protein molecule (lysozyme, see also Figure 1) exposed to 12 keV X-rays. Damage-induced error is plotted as function of pulse length and intensity.

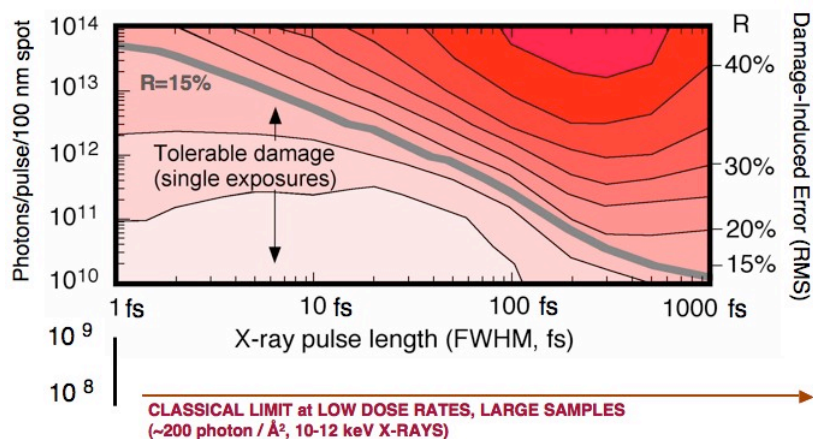


Figure 3: The landscape of damage tolerance as a function of pulse length and intensity. The figure shows contour plots of the expected damage-induced errors (R) in integrated diffraction patterns as a function of pulse duration and photon intensity for 12 keV X-rays. The weighted average R -factor describes the extent to which the information content of elastically scattered X-rays is degraded due to radiation damage during the exposure (0% = no change in the structure, 67% = structure randomised). We regard damage as acceptable if $R \leq 15\%$, and this value is shown as a grey contour line in the figure.

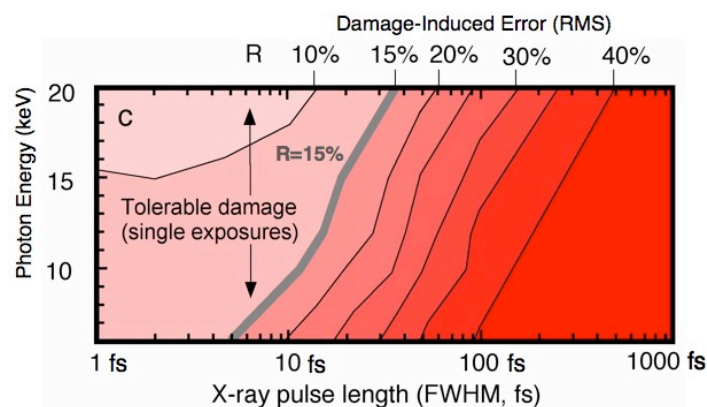


Figure 4: The landscape of damage tolerance as a function of X-ray energy and pulse length. The figure shows contour plots of expected damage-induced errors (R) as a function of X-ray energy when the total number of elastic scattering events per carbon atom is held constant ($I_{\text{tot}} = 1.33 \times 10^{12}$ for 6 keV; 1.85×10^{12} for 8 keV; 2.36×10^{12} for 10 keV; 3.0×10^{12} for 12 keV; 3.96×10^{12} for 15 keV; and 6.0×10^{12} for 20 keV X-ray photons/100 nm diameter). The weighted average R -factor describes the extent to which the information content of elastically scattered X-rays is degraded due to radiation damage during the exposure. We regard damage as acceptable if $R \leq 15\%$, and this value is shown as a grey contour line in the figure.

Figure 4 shows that data quality improves with increasing X-ray energies. This is due to a favourable change with wavelength of the ratios of elastic, inelastic and photoelectric cross sections in biologically relevant elements.

Figure 5 illustrates the calculated scattering intensities for a large macromolecule, Rubisco (Andersson, 1996), at various points along the grey contour line ($R = 15\%$) of Figure 3.

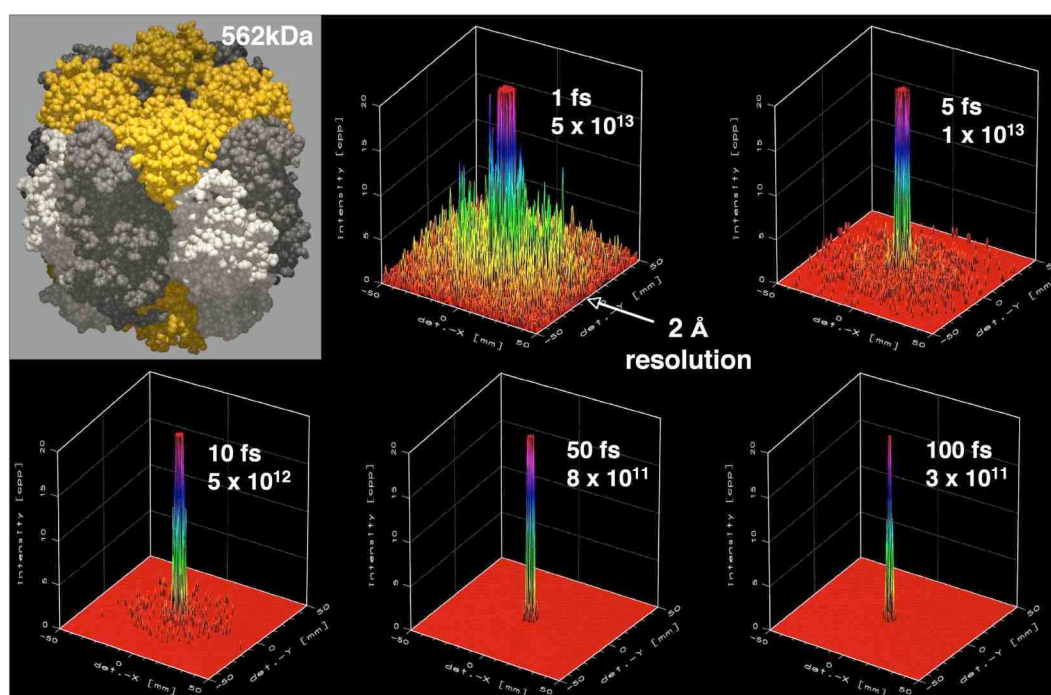


Figure 5. Expected scattering patterns for a large macromolecule, Rubisco, with 15% damage-induced error. The figure shows calculated diffraction patterns with 1, 5, 10, 50 and 100 fs long pulses, at pulse intensities at the corresponding points along the grey contour line ($R = 15\%$) of Figure 3.

The data show that very short and very intense pulses may permit atomic resolution imaging of this macromolecule. During such short and intense pulses (1-10 fs), numerous K-holes may be present at any one time, reducing the photoelectric cross sections of atoms in which they were produced, and thereby temporarily lowering the total number of primary ionisation events in the sample with X-rays. This effect can make the system radiation hardened to photoionisation during very short exposures. In addition, very short exposures can reduce Auger emission. If the pulse is shorter than the Auger decay time of the atom (i.e. shorter than around 10 fs for C, N, O), then significant Auger emission from these atoms will only happen after the passage of the X-ray pulse. This reduces damage during the exposure, as compared to longer pulses.

Achievable resolution as a function of pulse parameters and object size. A combination of results from the hydrodynamic continuum model [Hau-Riege et al., 2004] with the image classification model of Huidt et al. [2003] allows one to map out the landscape of imaging resolution, molecule size and pulse requirements [Hau-Riege et al, 2005]. The results are shown in Figure 6, which show that it will be possible to image single molecules at very high resolutions with very short pulse durations (atomic resolution with pulses less than about 5-10 fs).

First, the optimal photon energy for diffraction imaging was estimated by maximizing a figure of merit (FOM), defined as the ratio of signal minus noise to the radiation damage. As shown in Figure 6a, for pulses shorter than the Auger decay time (~ 10 fs for C), the optimum photon energy is 8 keV, and for longer pulses it is 13 keV, although the peak FOM is much smaller. Figure 6b shows the required x-ray fluence versus image resolution length and particle radius, required to achieve a large enough diffraction signal to classify the patterns. Figure 6c shows the pulse length requirements for x-ray imaging biological molecules with 12 keV photons, assuming no pre-orientation of the molecules. When the fluence requirements are relaxed by orienting molecules with laser fields, using nanocrystals containing only a small number of molecules, or helical molecules, or icosahedral virus particles up to 10-20 times longer pulses can be tolerated, see Figure 6d.

The physical models of the interaction of a small particle with the XFEL beam have led to insights

that offer additional ways to obtain high-resolution information with longer pulses. One such insight is that a tamper can be used to slow down the motion of atoms during the interaction with the pulse. The tamper may be a small water or helium drop that surrounds the molecule, and which has a total mass comparable to that of the molecule. Modern electrospray techniques can precisely control the amount of solvent left around the molecule and can be used to select an optimum layer thickness. Models show that as the molecule becomes charged, by the ejection of photoelectrons, the positive charge is confined to the surface, with a thickness of a Debye length. This layer is ejected first from the particle, and the Coulomb explosion proceeds from the outside in. A rarefaction wave propagates in from the surface at the sound speed, and hence the centre of the particle undergoes destruction later. The tamper is hence a sacrificial layer that preserves the structure of the molecule it contains. However, even with a tamper, the atoms at the centre of the particle quickly become ionized and are surrounded by hot free electrons. Since the X-rays scatter from electrons, the diffraction pattern is substantially modified by this effect. A strict figure of merit, such as used above, does not allow us to tolerate this effect. But since most of the charge is still localized around atoms, structural information is still present. It has been found [Hau-Riege & Timneanu, in preparation] that the diffraction pattern can be “repaired” to overcome the effects of ionization. The repair strategy assumes that the stoichiometry of the molecule is known and that the ionization occurs randomly and homogeneously. The method essentially filters the diffraction pattern to correct for the change in atomic scattering factors due to ionization. The cross terms, or the effect of different atoms being ionized differently is not corrected for, but it is found in simulations that the simple filtering has a dramatic effect and can correct for 90% of the damage. When coupled with the tamper, this leads to pulse durations about 10 times longer (i.e. on the order of 50 fs), as shown in Fig. 6 (e). Experiments with XFEL pulses are required to test these concepts.

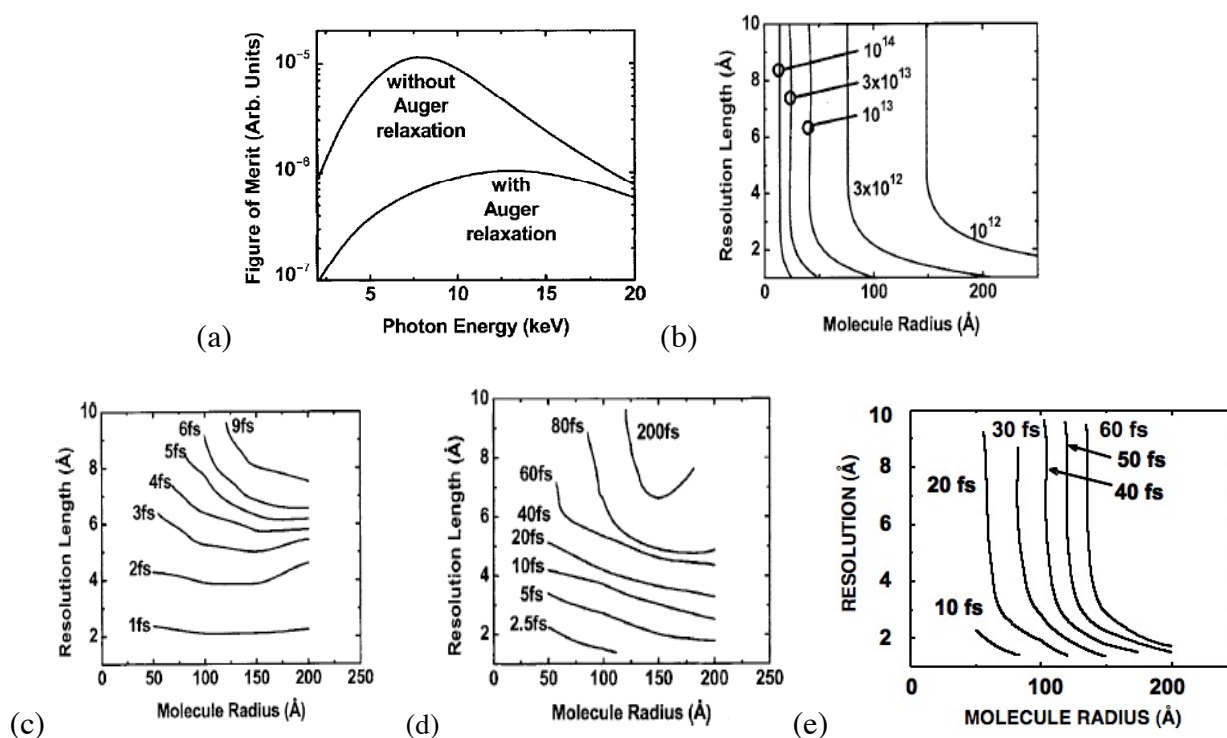


Figure 6. Resolution vs. radius for different X-ray fluences. (a) FOM for imaging conditions as a function of photon energy. (b) X-ray fluence requirements to classify two-dimensional diffraction patterns of biological molecules according to their orientation with 90% certainty. The curves are labeled with the X-ray fluence in units of photons in a 100 nm spot. (c) Plot of achievable resolution vs. molecule size for various pulse durations as limited by damage and classification. Atomic resolution imaging is achievable with pulse durations less than 5 fs and fluences greater than 10^{12} photons per 0.1 micron spot size. (d) Pulse duration requirements are significantly relaxed for samples that give 10 times larger scattering signal (e.g. viruses or nanocrystals). (e) The use of a tamper and diffraction pattern repair can give similarly dramatic increases in pulse length for single particles.

1.1.5 Finding image orientations, averaging, and building up a three-dimensional data set

Three-dimensional diffraction imaging will be performed by collecting a large number of noisy coherent X-ray diffraction patterns from a supply of identical particles. One diffraction pattern is collected per particle and the particles have random and unknown orientations. Image reconstruction requires (1) image orientation and assembly of a three-dimensional data set, (2) signal averaging to reduce the effects of photon shot noise, radiation damage and any other experimental noise source, and (3) phase retrieval. The methods of reconstructing a 3D image from a number of noisy diffraction patterns of random and unknown orientation are inspired by methods employed in single-particle electron cryo-microscopy [Frank, 1996; van Heel et al. 2000].

The greatest challenge is likely to lie in the signal-to-noise ratio of the diffraction images. The inherent power spectral density of the spatial distributions of matter lead to a rapid decrease of intensity with scattering angle (corresponding to increasing resolution). Molecule variability and incoherent scattering increase the noise, particularly at high scattering angles. At the same time, the incident number of photons should be kept at a minimum in order to reduce radiation damage. Therefore, it is crucial that the data set be redundant, and that we locate and average those redundancies to increase the signal-to-noise ratio.

Statistical studies have shown that a signal of less than one photon per pixel would be sufficient to correlate diffraction images of identical particles presenting the same view, assuming photon noise only [Huldt et al., 2003]. Correlation-based methods to average and orient large numbers of noisy, randomly oriented real-space images have been successfully developed in the electron microscopy community [Frank, 1996; van Heel et al., 2000]. Diffraction patterns are first classified into classes of like-orientation so that they can be averaged to increase the signal relative to noise [Huldt et al., 2003]. The average signal per diffraction pattern at the highest resolution, required for classification, is found to be much less than one photon per pixel, and an incident fluence of 10^8 ph/nm² is sufficient to achieve atomic resolution for particles greater than 15 nm radius [Huldt et al., 2003]. Averaged diffraction patterns must be oriented with respect to each other in 3D Fourier space, which may be achieved by the method of common lines (Figure 7), a technique widely used in electron microscopy, where the micrographs represent planar sections through the center of the molecular transform.

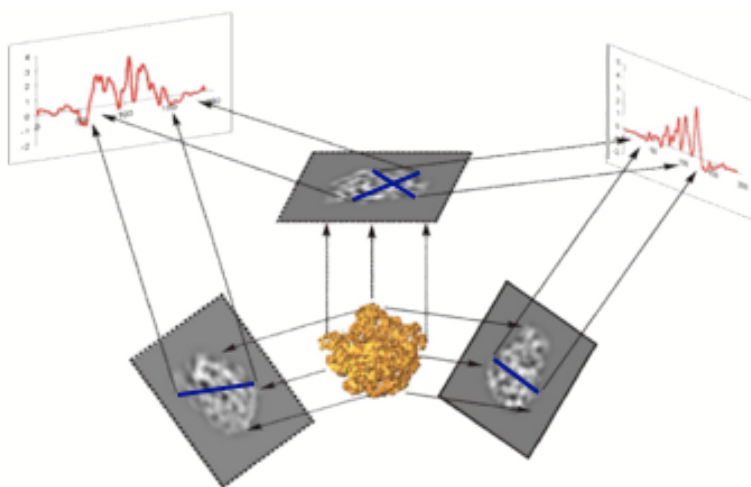


Figure 7. Three-dimensional reconstruction in electron microscopy/tomography [van Heel et al., 2000]. A 3D data set can be assembled from individual images based on the *common lines projection theorem*. The common line is a hinge axis in tomography.

Diffraction images are different and represent spherical sections. Each pair of images will intersect in an arc that also passes through the origin of the molecular transform (Figure 8).

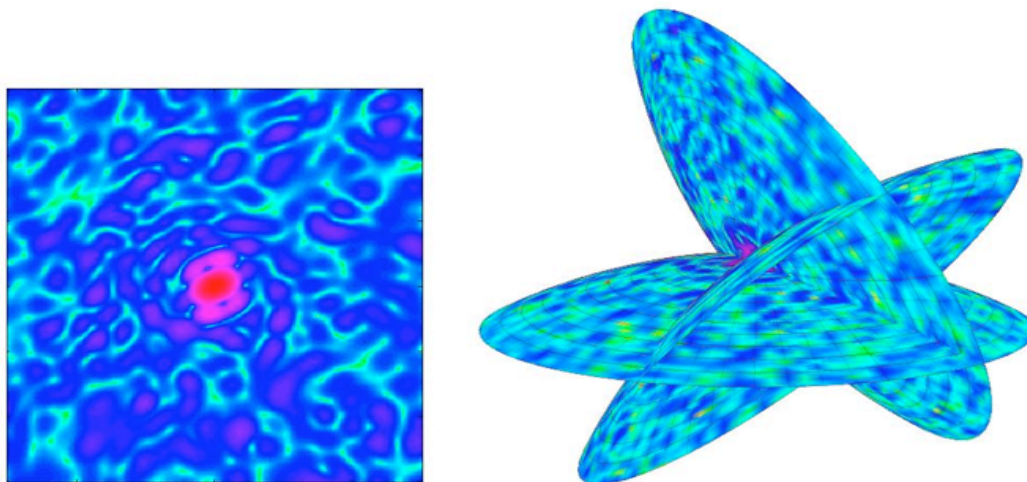
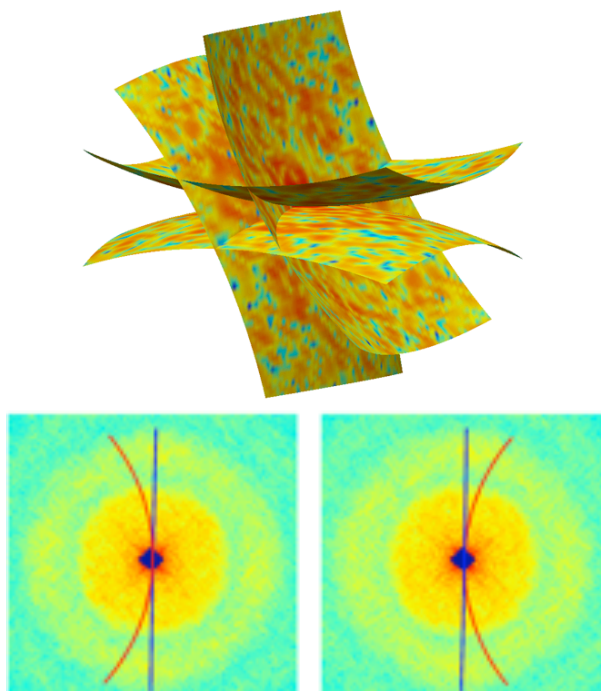


Figure 8. Intersection of diffraction images along common arcs in diffraction space. Figure on the left shows a predicted diffraction image for lysozyme. Figure on the right shows three different diffraction images of lysozyme intersecting along common arcs in diffraction space. Each arc gives a three-dimensional fix [Huldt et al. 2003].

If the signal is strong enough for the line of intersection to be found in two averaged images, it will then be possible to establish the relative orientation of these images. We note that due to the curvature of the sections (especially at X-ray wavelengths), the common arc will provide a three-dimensional fix rather than a hinge-axis. Moreover, the centric symmetry of the modulus of the molecular transform ensures that we obtain 2 independent repeats of the common lines in the two images. This feature provides redundancy for determining sample orientation, and is unique to diffraction images (Figure 9).



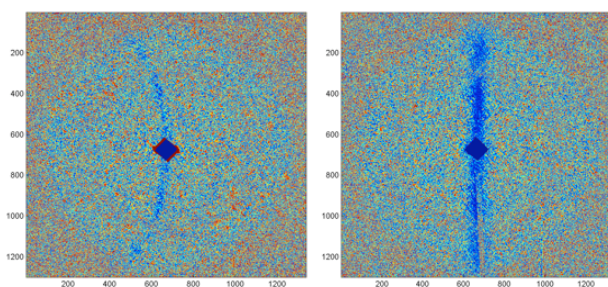


Figure 9 Intersection of two Ewald spheres with their centrosymmetric opposites. Centrosymmetry gives an extra intersect as there are *two* common arcs of intersection in *each* diffraction pattern (middle). The images in the middle show the expected arcs of intersections in two diffraction patterns from the experimental pyramid X-ray diffraction data set from Figure 10. Images at the bottom show these very lines of intersections when the experimentally obtained patterns are subtracted from each other pair wise (Huldt et al., in preparation).

1.1.6 Laser alignment of molecules

As a further possibility, it may be feasible to record 3D diffraction datasets of molecules using longer pulse durations and lower fluence by using a polarised laser to orient the molecule (or a number of molecules) and assist the data assembly. Diffraction data would be collected tomographically by rotating the polarisation of the laser once enough signal has been accommodated at a particular orientation. A linearly polarised AC laser field will induce a torque on a molecule that has an anisotropy in its polarisability (usually due to its non-spherical shape). Without damping (as is the case of molecules in vacuum) the molecule will oscillate with a period that depends on its moment of inertia, the polarizability anisotropy and laser intensity. This oscillation time is calculated to be 1 to 10 ns for small proteins to large complexes [Starodub et al., 2005]. If the rise time of the intensity experienced by the passing molecule is sufficiently slower than the oscillation period, alignment will be induced adiabatically and the molecule will orient along the direction of the electric field vector without oscillating [Larsen et al., 2001]. The rise in intensity can be achieved simply by shaping the beam intensity (a Gaussian profile will suffice) with a length scale that depends on the particle velocity, to achieve rise times of about 100 ns. The degree of alignment depends on thermal fluctuations. The equipartition theorem applied to a harmonic oscillator gives the result that the alignment error varies as the inverse of the square root of the laser power and as the square root of the molecule temperature. The rotational temperature can be brought down to a few Kelvin by the supersonic expansion of the molecule into vacuum that occurs in the injection process, which should result in alignments of 1° to 10° for proteins such as lysozyme under adiabatic conditions [Starodub et al., 2005]. This requires an intense laser field, on the order of 10^9 W/cm², which should be in the near infrared, far from vibrational resonances in the molecule.

The interaction energy of the molecule with the laser beam is the same for a molecule aligned parallel or antiparallel to the field axis and only a DC electrostatic field can break this symmetry. Nevertheless, recent work has shown that an image can be reconstructed from a diffraction pattern that is an average of these two directions [Elser, 2005]. Alignment of molecules along all three axes (with a sense ambiguity in each direction) has been achieved with high fields and simple inorganic molecules using elliptically polarized laser fields [Larsen et al, 2000].

The achievable image resolution will be limited by the degree of alignment, and in general for a length of the molecule of L and a standard deviation in the angle of orientation of $\Delta\theta$, the resolution will be about $L \Delta\theta / 2$. For small proteins, such as lysozyme, which are only $L = 4.5$ nm long, a misalignment of 10° gives a blurring of about 0.4 nm. More experimental studies are required to determine what degree of alignment is possible and to determine the required laser parameters.

Larger particles require much less laser intensity due to their larger polarizabilities (tobacco mosaic virus particles can be aligned in water with static fields [O’Konski and Zimm, 1950]). However, they must be aligned with far greater precision for a given image resolution, which is harder due to the rotational temperature. Thus for large single particles, which will give large diffraction signals, laser alignment may be used *for reducing the number of classes needed to classify a pattern*. Laser alignment will provide a useful platform for developing techniques of single-particle diffraction imaging using the characteristics of pulses in the first stages of XFEL development, allowing us to achieve scientifically relevant results and to better validate models of the interaction of molecules with XFEL pulses.

1.1.7 Methods for phasing

A number of methods exist for recovering phases for objects that have a finite size, or “support”. These include oversampling of continuous molecular transforms [Bates, 1982; Fienup, 1982; Sayre, 1990; Szöke 1999; Miao et al., 1999], holographic imaging methods [Szöke, 1986; 1993; Tegze and Faigel, 1991; 1996; Faigel and Tegze, 1999], holographic data evaluation methods [Szöke, 1993; 1997], classical methods of crystallography, and techniques for phase extension from lower resolution electron/X-ray cryo-microscopy images.

The past few years have seen the development of robust algorithms in solving the phase problem through oversampling the diffraction pattern, and this seems to be a most promising technique for the future. The 3D diffraction transform of a non-periodic particle is continuous. Only the diffraction amplitudes are sampled at discrete points by the pixellated detector and the process of classification. The measured diffraction intensities are proportional to the modulus squared of the Fourier transform of the wave exiting the object. On their own, these diffraction intensities are insufficient to back-transform to form an image in real space. That inversion requires knowledge of both the diffraction intensity and phase. If the diffraction pattern intensities are sampled finely enough, then it is possible to solve for the diffraction pattern phases [Bates, 1982; Fienup, 1982]. The solution to this non-linear inversion problem is usually obtained iteratively by sequentially enforcing known constraints in reciprocal space and in real space. Specifically, in real space we assert that the image has zero scattering strength outside the area of the object’s boundary (called its support) [Fienup, 1982], whilst in reciprocal space the squared modulus of the Fourier transform of the image must equal the measured diffraction intensities. Such algorithms have now been used successfully for image reconstruction in X-ray diffraction experiments [Miao, 1999; Robinson, 2001; He, 2003; Marchesini, 2003b; Williams, 2003; Miao, 2003; Shapiro, 2005; Chapman, 2006]. An example of a reconstructed 3D image is shown in Figure 10.

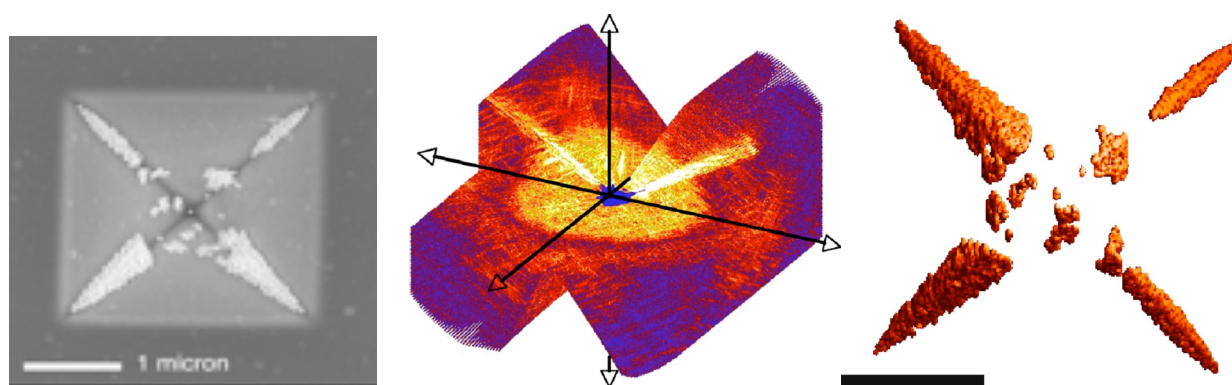


Figure 10. Coherent diffraction imaging and image reconstruction [Chapman et al., 2005]. Three-dimensional diffraction data (middle) recorded from a test object (left), consisting of 50-nm diameter gold balls on a silicon-nitride

pyramid-shaped membrane, at a wavelength of 1.6 nm, and a rendering of the *ab initio* 3D image (right) reconstructed from the diffraction intensities to a resolution of 10 nm. The diffraction data were obtained by rotating the specimen in 1° increments from -70° to +70°, and then interpolated onto a 1024³-element array. A quadrant of the diffraction dataset has been removed for visualization in the central rendering of the 3D diffraction intensities. The gold balls seen in the rendering of the 3D reconstructed image on the right fill the inside edges of the silicon-nitride pyramid. The scale bar is 1 micron.

The algorithms usually require that the support of the object be known *a priori*, and the closer the support to the actual object boundary, the better the reconstruction. The algorithm called SHRINKWRAP successively refines an estimate of the support from a current estimate of the image [Marchesini, 2003]. This algorithm does not require the support to be known and is remarkably robust at finding the smallest image support that contains the majority of the image intensity.

Another program that has been used successfully to phase experimental data is SPEDEN [Hau-Riege et al, 2004]. This uses a constrained conjugate gradient solver to find the amplitudes of 3D Gaussian blobs whose calculated diffraction intensities match the measurements while also minimizing cost functions based on constraints (including a low-resolution target, 2D projections or known phases). The algorithm finds the optimal image that fits all the constraints and, since it only ever performs calculations from real to Fourier space, it never needs to interpolate data onto a regular grid. However, as a local optimizer, it does not have as large a volume of convergence as the iterative transform algorithms and it may be used in XFEL imaging as a way to refine images produced by SHRINKWRAP and avoid artifacts due to missing data.

1.2 An overview of some of the possible experiments in structural biology

There is a general class of experiments that utilises the short time structure and the potential for very high intensities of the XFEL beam, and lead up to the exciting new regime in imaging outlined in the first sections.

It is immediately clear that small crystals, nanoclusters and two dimensional crystals can be studied at very high time resolution already with low photon intensities from the beginning. In the following sections we will outline some exciting biological problems that become doable this way. Note that the saturation of the beam plays only a secondary role in these experiments, but will be important later.

It was emphasised in the introduction that damage by the incident radiation is the ultimate limit to the resolution of imaging techniques in biological molecules. It is ~ 200 photons/Å² at 10 keV incident photon energy. The primary damage is caused by the ~ 10 keV photoelectrons and the ~ 250 eV Auger electrons that are absorbed in the sample, causing secondary ionisation and, eventually, chemical damage. It is accepted wisdom that the damage in electron microscopy is about 1000 times more benign with large samples than damage with X-rays. This would imply that single particle imaging techniques would yield 10 fold lower resolution by X-ray scattering. There are three mitigating factors in favor of X-rays. First, the scattering geometry of X-rays allows large angle data collection and the clean and simple geometry avoids distortions in the image. Second, if small particles can be used, the primary photoelectrons deposit only a small fraction of their energy and the damage should be ameliorated by ~ 20 - 40 fold. Third, time honored techniques of attaching heavy atoms to clusters should increase the signal to noise ratio of single particle images.

In order to utilise the advantages outlined in the previous paragraph, two additional developments are needed: focusing of the beam to about 0.1 micron, and the development of single particle injection techniques. The latter is well advanced in this partnership. We anticipate that very mildly

ionised droplets will be injected into the focused beam just in time. When mild focusing of the beam becomes available, e.g. the focusing of the projected 10^{12} photons in one pulse into a focal diameter of $1/2$ micron, giving a flux of 2×10^4 photons/ \AA^2 at the focus, single particle images can be obtained using averaging techniques similar to those used in electron microscopy. This should open the door to virus structures, including their genomes, to the study of membrane proteins, to very high time resolution studies of other, known structures.

In summary, we will argue below that new and exciting biological problems will be open for investigation from the first day the beam becomes available. We foresee that our interim technical efforts in sample handling, attaching proteins to viruses and developing algorithms for single particle imaging will fully complement developments with the source. The experiments outlined below represent major research lines far beyond the limits of currently available methodologies. We also hope that new research areas may emerge in a field as explosive as biology today by the time XFEL becomes available. Shortening the pulse length would be a key improvement in all planned applications, including those described in the biological proposal. With shorter pulses, tighter focusing, very fast detectors and development of accurate injection techniques, biomolecular imaging should reach the extreme regime outlined in the theoretical part of this proposal.

Nanocrystals are open periodic structures with submicron dimensions. All macroscopic crystals start as nano-crystals. X-ray lasers may offer completely new avenues for structural studies on nano-crystalline samples. No such studies are currently possible. When a crystal is small, the Bragg peaks are broadened and the intensity between the Bragg peaks is not negligible. Both the Bragg peaks and the intensity between the peaks carries structural information. The oversampled diffraction pattern visible between Bragg peaks can directly provide phase information.

Two-dimensional crystals of macromolecules. Various estimates show that the number of different membrane proteins in various genomes is similar to the number soluble proteins, yet there are only a small number of structures known today for integral membrane proteins as compared to well over 30,000 structures for soluble proteins. An understanding of the structure-function relationships in membrane proteins would make invaluable contributions to biochemistry, physiology and medicine, and would produce a substantial socio-economical impact (about 70% of all known drugs target membrane proteins).

Two-dimensional crystals of membrane proteins may be obtained from a number of membrane proteins, e.g. by epitaxial crystal growth methods. The intensity of the scattered X-rays even from a small two-dimensional array can be considerable with XFEL pulses. Integration of the diffraction rods requires images with different sample orientations to be recorded, and this could be achieved by merging data from several randomly oriented samples. This procedure could provide structural data on membrane proteins in bilayers.

Closed nanoclusters for structural studies. In contrast to nanocrystals (which are small but open periodic structures with translational symmetry), the nanoclusters we refer to here are closed structures, which may be periodic (like oligomeric proteins or the capsids of virus particles) but have no translational symmetry. Such nanoclusters come in very well defined sizes or can be made to make up well defined sizes and geometries. We are developing procedures for assembling proteins of choice into regular nanoclusters for subsequent structural studies. Methods are available for the specific attachment of target proteins onto the surface of regular templates, e.g. icosahedral virus capsids. These methods will be applied to structural studies on soluble and membrane proteins at free-electron lasers. Assembling protein molecules into nanoclusters will increase the intensity of scattered radiation from otherwise small proteins. We wish to use modified viruses to construct well defined nanoclusters of a number of different proteins, including membrane proteins.

Expected outcome: structures for "uncrystallisable" proteins attached to the surface of regular templates.

Virus structures and the structure of viral genomes. No high-resolution structure is available for any genome today. Small spherical viruses are among the simplest replicating systems in biology, yet the packing of the nucleic acid inside the capsid, and the factors affecting viral assembly, stability and disassembly are still not understood. Only a superficial picture is available today on the packing of the genetic material in intact viruses. This is due to the fact that in most viruses, the outer protein shell obeys the space group symmetry, while the inner material does not. As a consequence, the image of the otherwise tightly packed nucleic acid inside the virion is rotationally averaged over some angular range. From images collected, we propose reconstruction to recover the structure of the inner part (the genome) of the virus. Further experiments will focus on the assembly/disassembly of the virus. Of particular interest are viruses that cannot be crystallised (e.g. human immunodeficiency virus, herpes simplex virus, and many other viruses). X-ray studies on viral particles in the gas phase (or in vitreous ice) could open up ways to capture elusive intermediates within the "life cycle" of a virus. An understanding of the functional dynamics of viruses may offer a means to interfere with infection. Experiments will capture dynamic events, e.g. steps in the assembly and disassembly of the virus, and studies on key initial steps in a viral infection.

Structural studies on single protein molecules. The need for crystals for high-resolution structural studies is a serious limitation today. Currently, this excludes a very large proportion (>60%) of proteins from detailed structural determination, and hinders progress in the area of structural genomics. Structures accessible today for analysis do not represent a random selection of proteins, and knowledge gained on "crystallisable" structures may not automatically translate into knowledge about "non-crystallisable" structures. When sufficiently short pulses and a reasonable high pulse intensity will become available in the focus, studies on large single individual molecules (probably with molecular masses in excess of 100,000 dalton) may become routinely possible.

New horizons in time-resolved experiments. Biological function is a four-dimensional property. Time-resolved studies on structure, function and dynamics with X-ray lasers could cover catalysis, protein folding, nucleic acid folding, the assembly/disassembly of biomolecular complexes, viral uncoating, viral infections and so on. Such studies will first become possible on nanoclusters and nanocrystals as they require less stringent beam parameters than studies on single molecules. Later on, these experiments may be extended to individual biomolecules or cells.

Photochemical reactions. Marriage of femtosecond laser spectroscopy with femtosecond time-resolved structural studies will be possible.

Diffusion triggering, stop-flow studies. While certain key reactions in life are photochemical, most enzymes participate in diffusion-dominated processes with their reactants and partners. Time-resolved structural studies on diffusive processes in crystalline enzymes are difficult due to problems with mixing enzyme and reactant. Structural studies are only possible on intermediates which accumulate transiently in the crystal during a reaction. This requires a relatively fast binding followed by a relatively slow reaction. Due to the generally lower activity of crystalline enzymes, uniform catalysis can often be triggered by diffusing reagents (e.g. substrates) into crystals. However, the speed of diffusion and ligand binding sets an upper limit to the speed of reactions which can be analysed this way. Past results show, that in an average-sized protein crystal, half saturation binding with small ligands can be reached within about a minute. One obvious possibility for lowering diffusion barriers is to reduce the sample size, get rid of the crystal, and use *stopped-flow* type mixing techniques for time-resolved experiments, where the sample can be sprayed into the beam after certain "aging" times following rapid mixing. This can be done on purified proteins,

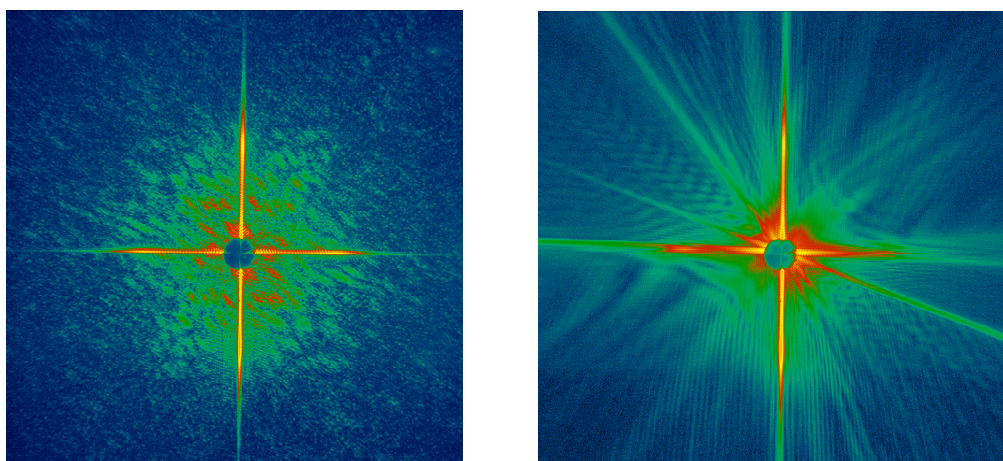
on nanoclusters, or on nanocrystalline slurries of enzymes instead of diffusing reactants into large single crystals. With very small samples, the vast majority of solution kinetic techniques and methodologies will suddenly become available for time-resolved structural investigations. We foresee that container-free sample handling methods based on spraying techniques, will open up new horizons here.

X-ray diffraction of whole cells. A challenging problem in cell biology involves the imaging of whole, eukaryotic cells at high resolution and with good resolution depth. X rays offer an opportunity for imaging whole, eukaryotic cells of 10 micrometer thickness at high resolution, and a free electron laser may be the only way to obtain the highest resolution data sets.

1.3 First results

1.3.1 Demonstration of flash-diffraction imaging with a soft X-ray free-electron laser (the *FLASH* facility at DESY)

The first experimental verification of the principle of coherent flash-diffraction imaging has recently been performed at the *FLASH* facility [Chapman et al. 2006: Ultrafast coherent diffraction imaging with a soft X-ray free-electron laser. Submitted]. The results show that an interpretable diffraction pattern can be obtained before the sample turns into a plasma when exposed to an intense 25 fs long photon pulse at 32 nm wavelength (Figure 11). In these experiments the beam was focused to a peak intensity of up to 10^{14} W/cm². We estimate that the absorbed energy density was approximately 20 eV/atom in the silicon nitride and that the material *reached a temperature of about 6×10^4 K before vaporizing*. Significantly, the image obtained by phase retrieval and inversion of the diffraction pattern shows no discernible sign of damage, and the object can be reconstructed to the resolution limit of the detector (62 nm with 32 nm photons). Damage occurs only after the pulse traverses the sample. A second exposure shows scattering from the hole that was created by the first pulse. These results provide the first experimental evidence for the basic principle of flash imaging, and this is the very technique we expect to lead to atomic resolution studies when hard X-ray FELs become available.



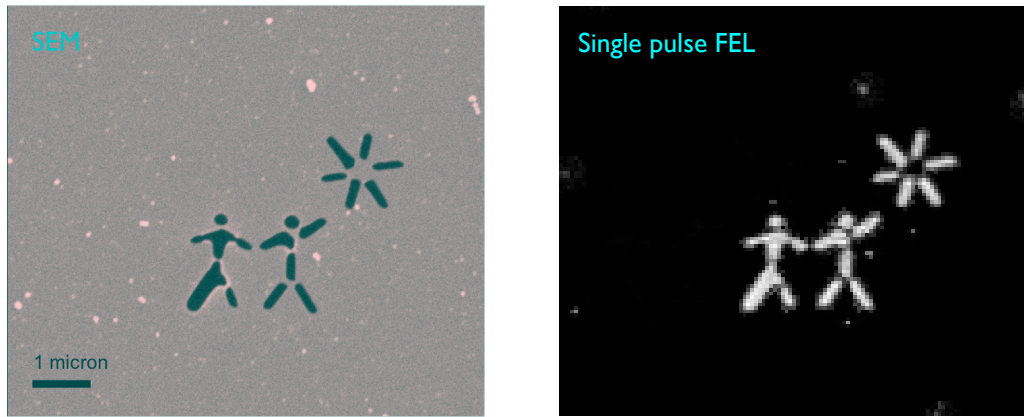


Figure 11. Demonstration of single-pulse coherent diffraction imaging at 32 nm wavelength (February 2006). *Top left:* A diffraction pattern recorded with a single 25 fs long FEL pulse at 32 nm wavelength from a test object placed in the 20 micron focus of BL2 at the *FLASH* facility of DESY (peak intensity: up to 10^{14} W/cm²). We estimate that the absorbed energy density is approximately 20 eV/atom in the silicon nitride and that the material reached a temperature of about 6×10^4 K before vaporizing. *Top right:* The diffraction pattern recorded on the second pulse some 20 seconds later, showing diffraction from the hole in the sample created by the first pulse. The sample was a pattern cut into a 20-nm thick silicon nitride membrane, shown at the *bottom left* from scanning electron microscopy (SEM). *Bottom right:* The image reconstructed to the resolution limit of the detector at 32 nm wavelength (corresponding to 62 nm resolution) from the single-shot diffraction pattern using the SHRINKWRAP phase retrieval algorithm. The algorithm only used the measured diffraction intensities and the knowledge that the diffraction pattern was over-sampled. We did not use the SEM image in the reconstruction. [Chapman et al. 2006; *Ultrafast coherent diffraction imaging with a soft X-ray free-electron laser*, submitted]

1.3.2 Ultra-fast transmittance and reflectance data from the *FLASH* free-electron laser

Other supporting data show that **transmission and reflectance** by solids remain *linear* up to 10^{14} W/cm² (the maximum that could be reached so far at the *FLASH* facility, absorbed energy in SiO₂: about 100 eV/atom), provided that this energy is deposited in the sample in a very short photon pulse (~ 30 femtoseconds at *FLASH*) [Sobierajski et al., in preparation; Juha et al. in preparation; Sokolowski-Tinten et al., in preparation]. The sample turns into a plasma, and it is destroyed *after* interaction with the intense photon pulse. Similar studies show that the reflectivity of mirror surfaces (single component silicon or graphite, and multilayer mirrors made of silicon and carbon) remain constant over a very wide intensity range when interacting with a very short pulse from the soft X-ray free-electron laser of DESY. These mirrors work perfectly *once* near normal beam geometry even in a focused FEL pulse at 10^{14} W/cm² at 32 nm wavelength [Sobierajski et al., in preparation; Juha et al. in preparation; Sokolowski-Tinten et al., in preparation; Hau-Riege et al, in preparation]. Angularly resolved reflection data show that the period of the Si-C multilayer mirror changed as little as 3 Å *during the pulse* [Hau-Riege et al, in preparation] but the surface was destroyed *after* interaction with the intense photon pulse (Figure 12), in line with our expectations.

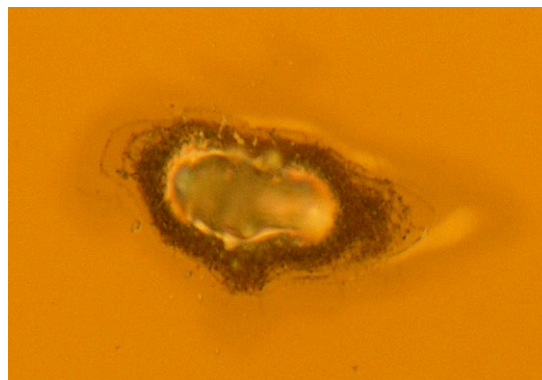


Figure 12. Nomarski photograph of a Si-C multilayer mirror after interaction with an intense pulse of the FLASH soft X-ray laser at DESY. Wavelength: 32 nm, focal spot size: ~20 micrometer, energy density: $\sim 10^{14}$ W/cm². The damage seen here developed *after* interaction with the intense photon pulse [Hau-Riege et al, in preparation; Juha et al., in preparation; Sobierajski et al. in preparation].

2. INSTRUMENT OVERVIEW

2.1 Design

The experiment consists of an apparatus to perform coherent X-ray diffraction imaging, and image reconstruction by phasing over-sampled diffraction patterns. The apparatus will include optics to focus the beam onto the sample to provide the necessary X-ray fluence together with pulse compression, as necessary. Samples will not survive the interaction with the focused XFEL beam, and so we can consider two fundamental classes of experiment:

- (i) three-dimensional diffraction imaging of reproducible structures,
- (ii) two-dimensional imaging of single objects (like small living cells), and
- (iii) three-dimensional diffraction imaging of single objects at low fluence.

Within these classes of experiment we can consider variations, such as simultaneous (or time delayed) stereo imaging of objects, imaging of reproducible structures in nanocrystals of different shapes, and imaging of two-dimensional membrane protein crystals. The third class listed above requires the unique coherence properties of the XFEL, and would be undertaken as a development step towards (ii) as well as providing learning that will be brought back to high-brightness third-generation sources.

Note that, while in a single diffraction pattern as recorded in scheme (ii) is two-dimensional, this information exists on the Ewald sphere in reciprocal space *and indeed does contain information of spatial frequencies in the depth direction*. This is manifested by the ability to computationally focus through the object, by numerically propagating the complex-valued wavefield that is the retrieved image [Chapman et al., 2005]. The depth resolution is simply the depth of field of the imaging system, given by λ/NA^2 , where NA is the numerical aperture of the detector. For $\lambda = 0.15$ nm, NA = 0.3, the depth resolution is 1.6 nm, compared with a transverse resolution of 0.5 nm.

Since the quantities of material of the sample under study will be minute, there should be very little other matter in the beam path, and therefore, the sample (and indeed the entire experimental apparatus) will be in ultra-high vacuum similar to conditions in conventional electron microscopy. The interaction chamber will house the sample manipulation and injection hardware, instrumentation for various diagnostics, and an in-vacuum area detector system to record the diffraction images (Figure 13).

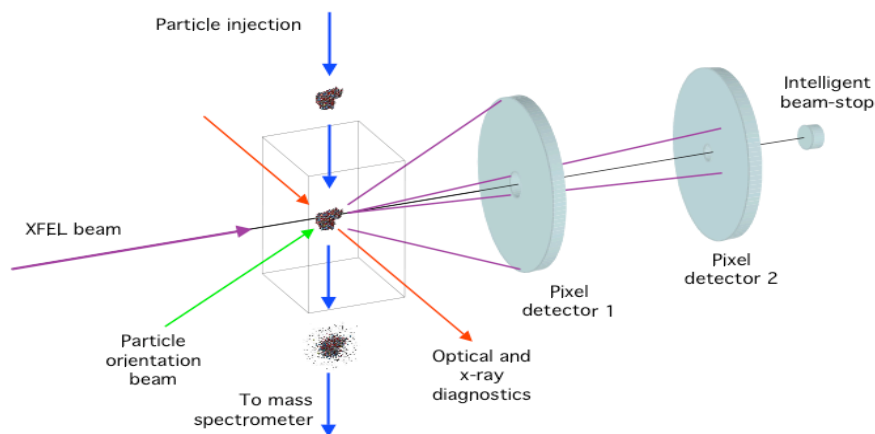


Figure 13. Schematic diagram of the single-particle diffraction imaging experiment.

Everything within the beam-path will contribute to the diffraction image. We will use **container-free** methods based on spraying techniques to select and rapidly inject single hydrated molecules, nanoclusters of molecules, viruses, and small living cells. Present sample injection and particle manipulation techniques will need to be improved significantly to achieve sufficient particle densities and injection precision. We will utilise **electrospray ionization (ESI) mass spectrometry** and related **ink-jet techniques** as methods for introducing clusters, particles, viruses and cells with well controlled properties into the gas phase [see, e.g. Siuzdak et al., 1996; Tito et al. 2000; Rostom et al., 2000; Ruotolo et al., 2005; Jayasinghe et al., 2006]. We will explore techniques for **trapping and aligning single particles** in the focus, using optical, electrostatic, or electromagnetic methods. In experiments on **solid samples**, the sample will be positioned and manipulated in the beam using a cryogenic goniostat adopted from electron cryo-microscopy. All experiments will require **shot-to-shot diagnostics** on intensity fluctuations, spectral profiles, pulse shapes and pulse lengths. Additional diagnostic data will be used to determine (a) if one or more of the particles were hit, (b) where were the hits along the beam path, and (b) how “good” were the individual hits. Such data can be obtained by measuring the UV-VIS light emission from the sample as it turns into a plasma, the electron and ion spectrum, and the fragmentation pattern at the end of the exposure. These data will be used in real time to veto bad shots.

Another approach will use **techniques from electron cryo-microscopy** (cryo-EM), and embed the sample into a thin layer of vitreous ice. These two approaches require different detectors, and have different focusing requirements. Micron sized objects, will primarily be studied with the cryo-EM sample holder, and experiments here will include studies on single cells, fibres (like Alzheimer fibres), organic and inorganic microstructures, micro- and nanocrystals, two dimensional arrays, etc. The station using spraying techniques will be used for experiments on nanometer sized objects, including single virus particles, macromolecular complexes, and single biomolecules.

Injected samples:

X-ray diffraction will occur in an ultra-high-vacuum chamber that will house sample manipulation and injection hardware, diagnostics, and two subsequent area detectors to record the high and low resolution parts of the diffraction pattern. Both of these detectors will have a small hole in their centre for the direct beam to pass through in order to (a) avoid placing any material (other than the sample itself) in the direct beam; and (b) allow the direct beam to go through the instrument for subsequent characterisation. Beam diagnostics will include the recording of spectral properties, intensity distribution, pulse duration, and wave-front characteristics for each pulse.

The design of the apparatus will incorporate experience with chambers developed for coherent X-ray diffraction imaging experiments at the *FLASH* facility at DESY and early experiments at LCLS when this machine becomes operational.

In most experiments, a diffraction pattern will be recorded in a single pulse and the sample subsequently destroyed. A large number of diffraction patterns will be recorded, from a supply of identical or equivalent samples, and stored on a computer to be processed into a three-dimensional (3D) diffraction dataset. This dataset will be processed (phased) to obtain a 3D image of the sample.

Pulse requirements - single shot imaging, samples injected into the gas phase:

X-ray Pulse and Pulse train:	Single pulse at 10-50 Hz rep. rate
Flux density per pulse	As high as possible
Pulse length	As short as possible (while maintaining high dose/pulse).
Wavelength:	15-1.0 Å and below (when this becomes possible)
Bandwidth	< 0.2 %
Transverse coherence	As high as possible
Higher harmonics	Useful for shorter wavelengths (creates less

damage/elastic event)	
Polarization (linear, circular)	Not relevant
Polarization (direction, hor/vert)	Not relevant
Focusing	Variable, down to 100 nm diameter
Pulse-to-pulse fluctuations	Not a factor (require diagnostics)
Shot to shot pointing stability	Not critical, but would prefer drift <10% of focus dia.

Pulse requirements - single shot diffraction from samples embedded in vitious ice:

X-ray Pulse and Pulse train:	Single pulse at 10-50 Hz rep. rate
Flux density per pulse	As high as possible
Pulse length	As short as possible (while maintaining high dose/pulse).
Wavelength:	15-1.0 Å and below (when this becomes possible)
Bandwidth	< 0.2 %
Transverse coherence	As high as possible
Higher harmonics	Useful for shorter wavelengths studies
Polarization (linear, circular)	Not relevant
Polarization(direction, hor/vert)	Not relevant
Focusing	Variable, down to 1 micrometer diameter
Pulse-to-pulse fluctuations	Not a factor (require diagnostics)
Shot to shot pointing stability	Critical, need minimal drift: <10% of focus dia.

Pulse requirements - diffraction tomography, samples embedded in vitious ice:

X-ray Pulse/Pulse train:	Pulse trains can be used (attenuated), 10-50 Hz rep. rate
Flux density per pulse	Low
Pulse length	Not a key factor at present.
Wavelength:	15-1.0 Å and below (when this becomes possible)
Higher harmonics	Useful for shorter wavelengths studies
Bandwidth	< 0.2 %
Transverse coherence	As high as possible
Polarization (linear, circular)	Not relevant
Polarization(direction, hor/vert)	Not relevant
Focusing	Variable between 20-1 micrometer diameter
Pulse-to-pulse fluctuations	Not a factor (require diagnostics)
Shot to shot pointing stability	Critical, need minimal drift: <10% of focus dia.
Pump-probe	A possibility
Synchronization requirement:	Around 200 fs (require diagnostics to better than this)

Components of the experiment:

(Xenon gas) Attenuator

Range $\times 10^6$
Steps continuous between 1-1000
Precision 20%

For 8 keV imaging experiments, the attenuation needs are to be able to align and characterize the experiment without destroying test samples. For the tightest focus (0.1 micron), that is about 10^6 attenuation (10^4 for 1 micron spot). This is assuming a sample with elements lighter than Si and 2 mJ pulses (and reducing to 10% of melt in the case of solid attenuators). It would be preferable for all attenuation to be done as far away as possible from the experiment. Far away, high (and not so high) spatial frequencies in the solid would have scattered out of the beam, and close they would not have propagated enough to redistribute energy.

Pulse energy precision of about 1%, but should be able to have good precision over the range of pulse energies given by the attenuator (10% precision OK at the highest attenuations). The 1% comes from the need to scale diffraction.

Photon energy: It would be good to know photon energy to 1 part in 1000 at least.

Monochromator (not needed for imaging single particles or biomolecules, but optional for micron-size samples):

Range 5 keV-24 keV

Resolving power 3000 to 10000

Focusing option 1 (20-1 micron):

Average: 5 microns x 5 microns

Included energy 80%

Position stability 0.5 microns

Focusing option 2

0.1 micron x 0.1 micron

Included energy 80%

Position stability 0.1 microns

Diagnostics

Flux (+/- 5%)

Pulse length (need to know to better than 10-30 fs resolution)

Wavelength (need to know to better than $10^{-3} \Delta\lambda/\lambda$)

Focusing optic needs: measure focal spot size and included energy

Spot size resolution 0.5 microns

Included energy 10%

Measure pulse shape and wavelength distribution

Measure wavefront of the pulse behind the experiment

2d detector

Pixel size smaller than 100 x 100 microns

Minimum 1000 x 1000 pixels

Dynamic range minimum 1000

Single photon detection

Detector noise < 0.06 of a single photon

Read out rate: faster than the repetition rate (min 20 Hz)

Modulation transfer function: >70% for modulations of period $4p$, where p is the pixel pitch.

Hole in detector < 4 x 4 pixels

Working environment

High vacuum (below 10^{-6} mbar, ideally 10^{-8} mbar)

Optics

The beam from the XFEL must be focused to a variable spot diameter 0.1 to 10 microns (variable), depending on the overall sample size and desired X-ray fluence at the sample (some experiments can use larger diameter beams, including the unfocussed beam). This requirement could be achieved by a Kirkpatrick-Baez mirror pair, or a grazing ellipsoidal mirror. No prefocusing optic would be used, and ideally the number of beam-directing mirrors should be minimized. A given

required spot size dictates a demagnification factor of the optic, which forms a demagnified image of the source. The optic to focal spot distance is equal to the source to optic distance multiplied by the demagnification, and so larger working distances can be achieved by placing the experiment further from the source. This has the added advantage of allowing larger aperture optics (since the unfocussed beam is bigger) and hence reduced fluence on the optical surfaces. The pulse duration will eventually need to be shortened, which could be achieved either by source modifications or by X-ray pulse compression, using, e.g. strained crystal diffraction (Chapman and Nugent, 2002) or asymmetric multilayer gratings. Our current estimates are that shorter than 50 fs pulses with more than 10^{11} photons/pulse are required for realistic single-particle imaging.

Other optical elements include apertures placed in the beam to block unwanted scatter from the beam line. These apertures will be larger than the direct beam. Also called guard slits, these have to be positioned and manufactured very carefully to prevent rather than contribute to the problem. Slits with “soft” edges can be made from wedges of perfect Si crystals (as one particular option). The innermost edge is transparent, slowly becoming opaque. This apodizes the diffraction pattern by removing high-spatial frequency from the slit structure. With coherent beams, several slits can be placed so that their edges are position in the node of the weak scattering pattern formed by the upstream edge.

Channel optics for detector shielding

We may find it necessary to shield the detector from scattering from gas or objects upstream from the sample, even with the judicious use of the apertures mentioned above. Ideally the detector would only be sensitive to rays travelling in directions emanating from the sample position, so that it would not detect scatter from any other object in the beam. Since downstream from the sample and off the axis of the direct beam the X-ray intensity is very weak (the scattering from the sample) it may be feasible to use structures such as radial Soller slits (or, more precisely, square-pore capillary arrays) accept only the diffracted light from a volume surrounding the sample (Figure 14).

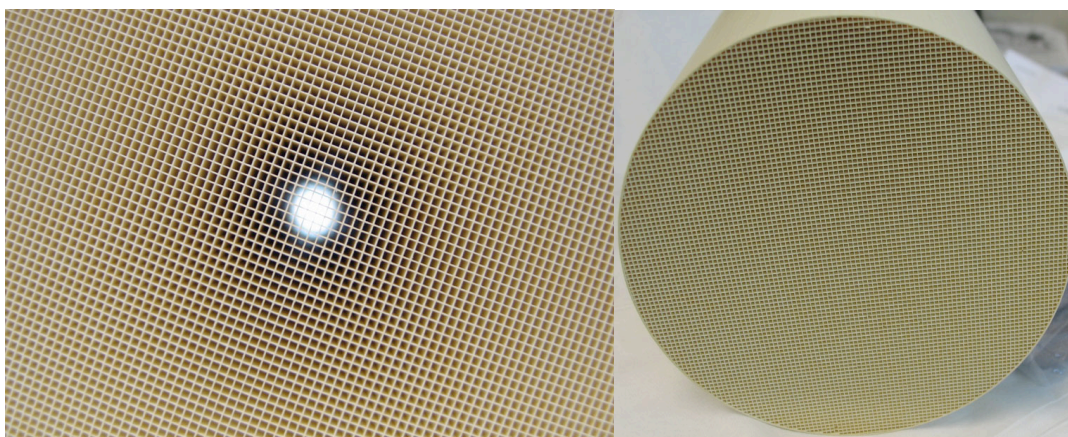


Figure 14. Glass Soller collimator for filtering out stray light at the detector. These can be manufactured to conical shapes, and can have tapered channels with dimensions down to about 10 micrometer x 10 micrometer.

Sample Handling

The main complexity and scientific challenge of the instrument will be in sample introduction and control. Since the quantities of material of the sample under study will be minute, there should be very little other matter in the beam. When imaging single molecules, the sample cannot be held on a substrate, since scattering from the atoms of the substrate will overwhelm the signal of the molecule itself, even if the substrate was a perfect Si crystal, for example. Therefore, the sample (and indeed the entire optics and experimental apparatus) must be at ultra-high vacuum (pressure $\sim 10^{-8}$ mbar, similar to cryo-EM).

CRYOMICROSCOPIC TECHNIQUES could be suitable for studies where a controlled sample rotation is necessary, e.g. in coherent diffraction experiments on intact cells and cell organelles using the unfocused, coherent beam of the X-ray laser. Electron cryo-microscopy performs structural studies on hydrated samples at low temperatures in a high vacuum environment. Existing EM equipment can be adopted for similar studies in an X-ray beam. Sample molecules and particles may be embedded in a thin layer of vitreous ice of a few hundred Å in thickness. Prior to an exposure, the sample(s) of interest may be located by UV/VIS fluorescence techniques, and once found, moved into the path of the X-ray pulse, using programmable positioning devices. Such a technique requires excellent pointing stability from the XFEL, with minimal pulse to pulse creep. If this can be achieved, the complete repertoire of methods in electron cryo-microscopy will become available for X-ray experiments. Vitreous ice surrounding the sample will contribute to background. This method may, however, be the method of choice for large samples.

SPRAYING TECHNIQUES: Particles, such as macromolecules or virus particles, will be injected from the outside into the X-ray beam in such a way that single particles intersect with the brief XFEL pulses. Ideally, one fresh, single particle is injected into every focused pulse at the pulse rate of the X-ray laser. To achieve this, the trajectories of the particles must be controlled both in space (< 1 micron) and time (< 10 ns), so that each one of them will be well aligned with the focused X-ray pulse. Present sample injection and particle manipulation techniques need to be significantly refined in order to position individual particles with sufficient precision. This approach would also require very high pointing stability from the XFEL beam.

Initial experiments will use clouds of particles without stringent requirements, achievable with current methods, with and without alignment with a polarised laser. These will be improved upon, by first injecting short, concentrated bursts of particles into the beam focus area and relying on statistical positioning of individual particles. Such experiments require a focal spindle of about 5 mm length, shot-to-shot diagnostics to determine whether a particle was indeed hit by the X-ray beam, and where did the interaction happen along the focal spindle. Such diagnostics could be provided, e.g., by fluorescence detectors and by a mass spectrometer that analyzes the fragments of the particle after the shot passes through.

More advanced techniques of particle introduction and manipulation include the injection of a few or even single particle into the beam at the proper time with well-controlled velocity, or trapping single particles at the XFEL beam focus using optical, electrostatic, or electromagnetic methods. This area of the project will need R&D prior to user operation. Electrospray ionization (ESI) and related ink-jet spraying techniques will be used for introducing samples (like molecules or single particles, such as viruses) into the gas phase.

Spraying techniques have been refined in recent years for their application in mass spectrometry of large proteins, supramolecular complexes, such as intact ribosomes [Rostom, et al., 2000], and even whole viruses [Tito, et al., 2000]. The charge imparted onto a particle by the ESI process is convenient for manipulating the particles in the gas phase by electrostatic forces. If necessary, a charge reduction scheme based on the charge-reduction electrospray method can be used to reduce the charge on electrosprayed molecular ions or particles to one or few elementary charges in a controlled way. For the particle introduction into ultra-high vacuum, we will expand on the aerodynamic lens or nozzle techniques used for single-particle mass spectrometry and bioaerosol mass spectrometry developed at LLNL. We will also explore techniques based on reverse micelles or helium droplets to provide a protective coat for the sample molecules if necessary.

As discussed in Sec. 1.1.6, alignment of single particles or clouds of particles will be performed

using a polarized near-infrared laser, focused to an intensity of about 10^9 W/cm². For adiabatic alignment in the non-viscous environment of a pure vacuum, pulse durations longer than 100 ns are required. If the beam is focused down to 10 micron, this requires a laser with pulse energies of more than 100 μ J, or alternatively a 1 kW continuous laser.

For initial experiments, large particles such as nanocrystals of membrane-bound proteins or other materials, and cells, will be supported in vitreous ice and manipulated directly into the beam by visual microscopy. The handling of these types of samples will be upgraded to be completely containerless by using a simple electrostatic system or special in-vacuum laser tweezers. This method would be ideal for diffraction imaging of membrane protein nanocrystals, but must be automated to enable collection of diffraction from thousands of individual crystals.

Detectors:

The diffraction pattern will be recorded on a pixellated detector subtending a solid angle dependent on the desired resolution, and a hole in the middle to avoid the direct beam. There must be sufficient pixels in the detector to over-sample the diffraction pattern, which depends on the sample size and desired resolution, as described below. The scattering from the sample covers a large dynamic range: it is strong very close to the central core, and at high angles there will be much less than one photon per pixel. Since the technique relies upon classifying and averaging a large number of patterns, the read noise must be considerably less than the photon count per pixel averaged over these patterns. Estimates of the noise level and dynamic range are given below, after first listing the requirements of pixel count and sampling. For larger structures, e.g. virus particles, and single cells, a finer sampling will be needed than for smaller objects (Figure 15).

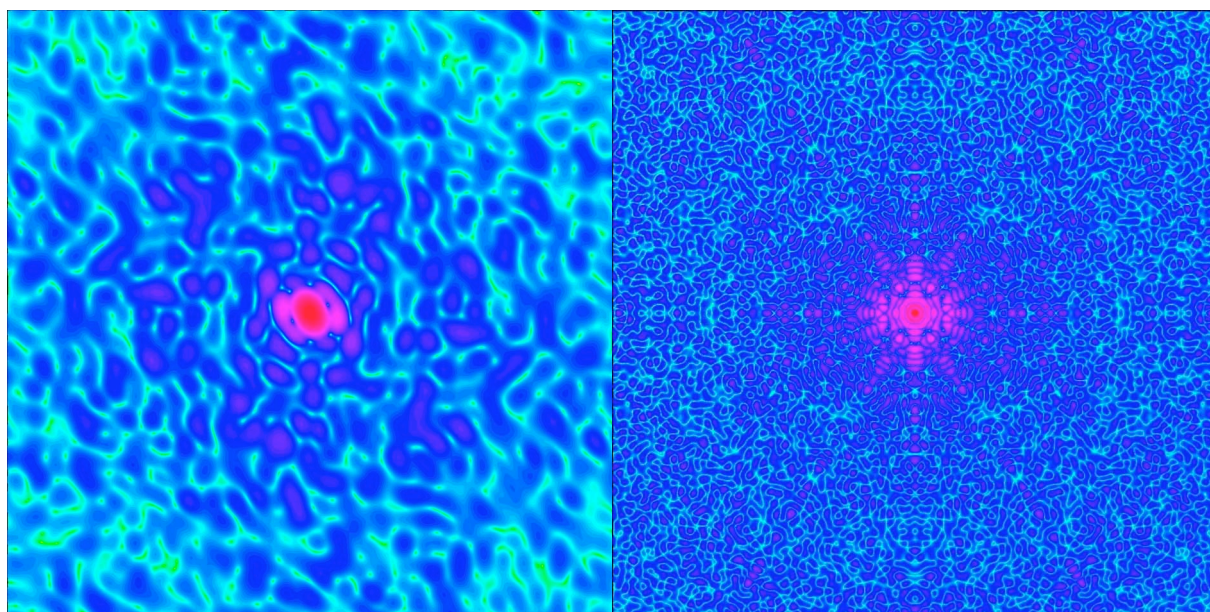


Figure 15. Planar section through the centre of the molecular transform of a small protein molecule (lysozyme, left) and of a larger virus capsid (tomato bushy stunt virus, right) at similar maximal resolutions. The level of detail is significantly different in the two pictures, and detectors should be able to resolve details in the patterns of even larger objects than a virus particle.

The detector size should not be larger than approximately 100 mm x 100 mm, in order to reduce the beam path from sample to detector. At better vacuum levels, this distance could be increased without increasing the background. Other desired parameters are a read-out speed per frame matching the pulse rate of the X-ray laser; a dynamic range of $>10^6$ for the entire pattern (for large single particles such as viruses); a dynamic range locally of ~ 1000 (a dynamic range of 10^{6-8} could

be achievable with two detectors, each with smaller range where the response of the second detector measuring the strong forward scattering component is reduced by appropriate means). Diffraction data may be supplemented by a lower-resolution image of the sample obtained with a *zone plate* or a *wave-front sensor* as it could prove valuable in enhancing the robustness of the oversampling phasing algorithms.

Pixel requirements

The pixel requirements simply depend on the number of resolution elements to sample the object of a given size at a given resolution, as described in Huidt et al. [2003], for example. To record to a resolution $f_{\max} = 1/d$ requires a maximum scattering angle 2θ given by $\sin \theta = \lambda f_{\max}/2$. For an object of finite extent of width D , its molecular transform (in reciprocal space) is band limited. The Nyquist sampling rate of the transform is $1/D$ in each dimension. To measure this transform to a resolution $1/d$, in one dimension, requires samples from $-1/d$ to $+1/d$ or $2D/d$ samples. In real space this corresponds to samples at intervals $\Delta x = d/2$, which is the largest sufficient interval to measure periods larger than d . The detector measures the diffraction intensities, which are the modulus squared of the molecular transform, or equivalently, the Fourier transform of the object's autocorrelation function. For an object of extent D the extent of its autocorrelation is $2D$, which means that the diffraction intensities are band limited with a Nyquist rate of $1/(2D)$. The phase retrieval algorithms do not necessarily require sampling at this rate but experimental experience shows better results with higher sampling. Note that sampling at a higher rate than $1/(2D)$ does not add any information to the measurement, but may improve the signal to noise ratio of the measurement. However, pixellated detectors do not sample at points but integrate over the active area of the pixels. This corresponds to a Modulation Transfer Function (MTF) that may decrease to zero at spatial frequencies (at the detector) of period $2p$, where p is the pixel width. The effect of the MTF is to apply an envelop to the reconstructed real-space image, which should be no less than 0.7 at the largest radial extent of the object. As such, the detector's MTF influences the required pixel count. The number of pixels along the width of the detector is given by $N = 2 D s / d$, where s is a sampling ratio per dimension (relative to the molecular transform Nyquist rate), with $s = 2$ in the case of maximum required sampling (for which the 0.7 MTF level should occur for pixel frequencies no lower than $1/(4p)$).

We estimate the maximum requirement for number of pixels is $N = 2000$, which corresponds to a particle size of 100 nm at a resolution of $1/(0.3 \text{ nm})$ and a sampling ratio of $s = 3$, or a particle size of 200 nm at the same resolution and a sampling ratio of $s = 1.5$. The larger sampling would be required if the detector MTF at $1/(2p)$ is about 50%. These are likely parameters for imaging of nanoparticles, and for the imaging of arrays of biological particles. For the cow-pea mosaic virus (CPMV) test object described below, which has $D = 32 \text{ nm}$, we require $N = 450$ pixels for $s = 2$ and a resolution of $1/(0.3 \text{ nm})$. This reduced pixel count will be sufficient for most small biological samples, and a larger pixel count detector will be needed for larger objects (Figure 15).

Image reconstruction can be achieved with considerable missing data due to a "beamstop" or a corresponding hole in the middle of the detector. However, the larger this region, the larger the uncertainty of various components of the image, and the less quantitative the image. If the central blank area covers no more than the central speckle, then the only missing data is essentially F_{000} . The speckle size for an object of width D is $1/D$ or $2s$ pixels in width. The dynamic range values given below were based on patterns with the central $2s \times 2s$ pixels excluded.

Simulations

Detector signals were computed for a test sample of a cow-pea mosaic virus (CPMV), with labeling of nanoclusters of gold [Wang et al., 2002]. This may be an early test sample for LCLS and XFEL

experiments. The atomic coordinates of the virus capsid (1NY7) were obtained from the EMBL-EBI Macromolecular Structure Database (<http://pqs.ebi.ac.uk/pqs-bin/macmol.pl?filename=1ny7>). The capsid structure is hollow (since the DNA structure inside is unknown), and this was filled in with carbon atoms in random locations and average density of 1.3 g/cm³ (less mass than DNA). When the gold labeling was applied, clusters were attached to the 65 symmetry sites (CYS 295) as described by Wang et al. [2002]. Each gold cluster was spherical with a diameter of 1.4 nm diameter, and contained 82 gold atoms (density of 18.8 g/cm³). The gold increases the total scattered photons by less than 10%. The CPMV has a diameter of 32 nm, and a total molecular mass of 14.8 MDalton (13.7 MDalton without the gold labeling). Larger sized samples have been simulated by arraying the CPMV structure in ordered and disordered groups of particles. The incident beam was modelled as a single-mode coherent Gaussian beam, with a waist diameter of 0.1 or 0.2 micron and total flux of 10¹² photons. The diffraction patterns were computed for a wavelength of 0.15 nm, with a program that computes the scattered intensity from a collection of atoms illuminated by a focused single Gaussian mode, in the Born approximation (and no atom motion or ionization):

$$I(\mathbf{k}_{out}) = \Omega r_e^2 I_0 \left| \sum_j f_j(\mathbf{q}) u(\mathbf{x}_j) \exp\{i\mathbf{k}_{out} \cdot \mathbf{x}_j\} \right|^2$$

where f_j is the structure factor for the j^{th} atom, r_e the electron radius, Ω the solid angle of a pixel, and an incident field described by Siegman [1986]

$$u(\mathbf{x}) = u(\mathbf{r}; z) = \sqrt{\frac{2}{\pi}} \frac{1}{w(z)} \exp\{-ikz + i\psi(z)\} \exp\{-r^2/w^2(z)\} \exp\{-ikr^2/(2R(z))\}$$

with

$$w(z) = w_0 \sqrt{1 + (z/z_R)^2};$$

$$R(z) = z + z_R^2/z; \quad \psi(z) = \tan^{-1}(z/z_R).$$

The parameter w_0 is the waist radius (radius at which intensity is $1/e^2$), and $z_R = kw_0^2/2$ is the Rayleigh range. The length $R(z)$ is the radius of curvature of the wavefront at a distance z , along the propagation axis, from the waist. The wave-vector magnitude is defined as $k = 2\pi/\lambda$, and the momentum transfer \mathbf{q} is given by $\mathbf{q} = \mathbf{k}_{out} - \mathbf{k}_{in}$ with

$$\mathbf{k}_{in} = [\mathbf{x}_j, -R(z_j)] k / \sqrt{\mathbf{x}_j^2 + R^2(z_j)}.$$

Note that $\int_0^\infty 2\pi r u^2(r; z) dr = 1$, and thus the parameter I_0 is the total number of photons in the beam. In the simulations the intensity I is quantized and photon noise added (normally distributed with standard deviation \sqrt{I}).

Signal

Results of simulations are shown in Figure 16. For these calculations, the centre of the particle was always positioned in the centre of the beam, at the z location of minimum waist ($z = 0$). For the calculations shown here we used $I_0 = 10^{12}$ and $w_0 = 0.1 \mu\text{m}$ (0.2 μm diameter waist) and $w_0 = 0.05 \mu\text{m}$ (0.1 μm diameter waist). The calculations were carried out for an X-ray wavelength of 0.15 nm. The simulated array size was 422x422 pixels, with $s = 2$, corresponding to a resolution of 1/(0.3 nm). Most of the pixel values are zero or one photon, and the maximum and total photon counts (excluding the central 4x4 pixels) was 1530 and 7.3×10^4 , respectively, for the 0.2 μm

diameter waist and 5540 and 3.0×10^5 , respectively, for the 0.1 μm diameter waist.

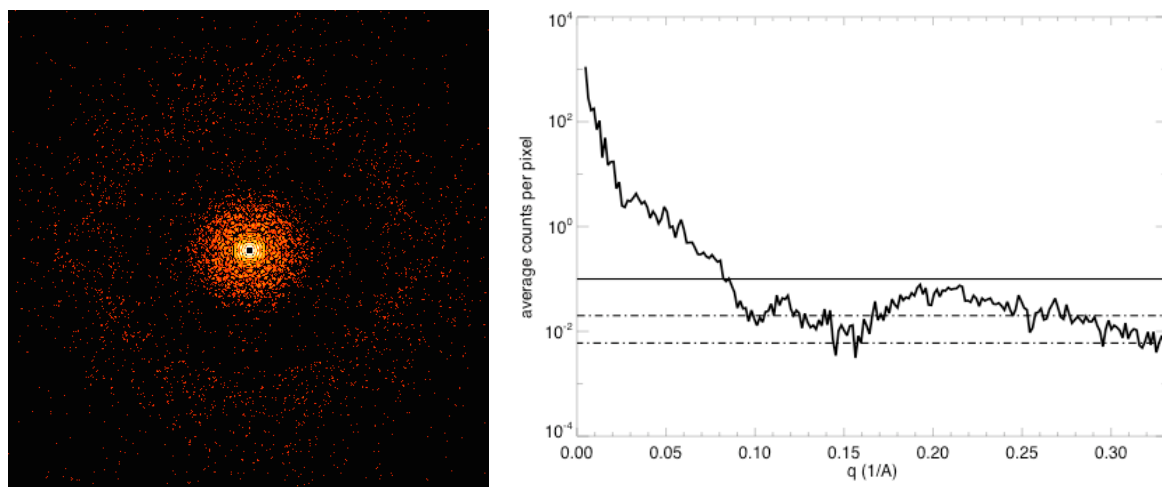


Figure 16: Simulated diffraction data from the cow-pea mosaic virus (CPMV) test object, for a beam waist diameter of 0.2 μm . The total photons in the incident beam was 10^{12} . The array is 422x422 pixels, corresponding to $s=2$, and a resolution of $1/(0.3 \text{ nm})$. The central 4x4 pixels were blocked. The total integrated photons is 7.3×10^4 . The intensities are displayed on a logarithmic greyscale. Away from the centre most pixels have one or no photons. The maximum photon count is 1530. The plot shows the radial average of the photon counts. The solid line is at 0.1 counts, the minimum counts needed to classify [Huldt et al., 2003], and the dotted lines denote the noise level for 10 and 100 averages, for a detector noise of 0.06.

Note that larger samples do not necessarily give larger signals; there are a fixed number of photons per pulse and samples larger than the beam will require a proportionally larger beam and hence lower fluence. Larger signals will be achieved with thicker objects and objects of higher-Z, and as such the experiments of nanoscale inorganic samples will give stronger signals. Larger signals will also be achieved with arrays (2D or 3D) of particles, due to coherent addition in the Bragg directions. Similarly, the imaging of large objects at low resolution (e.g. single-shot imaging of micrometer-sized cells beyond the radiation damage limit) will produce larger photons per pixel due to the coherent addition, in the forward direction, of scattering from atoms within a single resolution voxel [Sayre and Chapman, 1995]. For the case of crystals and arrays of identical unit cells, the photon count will increase in the Bragg peaks by a factor n^2 , where n is equal to the total number of unit cells illuminated. The signal between the Bragg peaks will only increase in proportion to n , but this signal can be built up by averaging, once classification has been achieved on the Bragg peaks. For a 2D crystal with n unit cells, if the beam size matches the object size then the incident fluence will be proportional to $1/n$ and the Bragg peaks will increase in photon count by n , not n^2 . For the simulations we have run on 2D and 3D crystals of 5×5 and $5 \times 5 \times 5$ unit cells, we typically see a factor <10 increase in photon count. In all these cases, the pattern is strongest near the zero frequency and locally (within a 10×10 pixel patch) the intensity changes by a factor of about 1000.

Noise

Datasets will be assembled by classifying patterns into classes of like-orientation and averaging patterns within each class. Huldt et al. [2003] showed that accurate classification could be performed with as little as 0.1 photon counts on average per pixel, at the highest resolution of the pattern. For the CPMV particle simulation, this means we can classify out to the full resolution of the simulation, of $1/(0.3 \text{ nm})$, for the case of a 0.1 μm diameter waist (Figure 17). This requires that when we sum together ten diffraction patterns the accumulated noise must still be less than one photon. Choosing a noise level of 0.2 photon (SNR = 5, Rose criterion) in the ten-frame sum, the noise per pixel for each pattern should be no larger than $0.2/\sqrt{10} \approx 0.06$ photons per pixel.

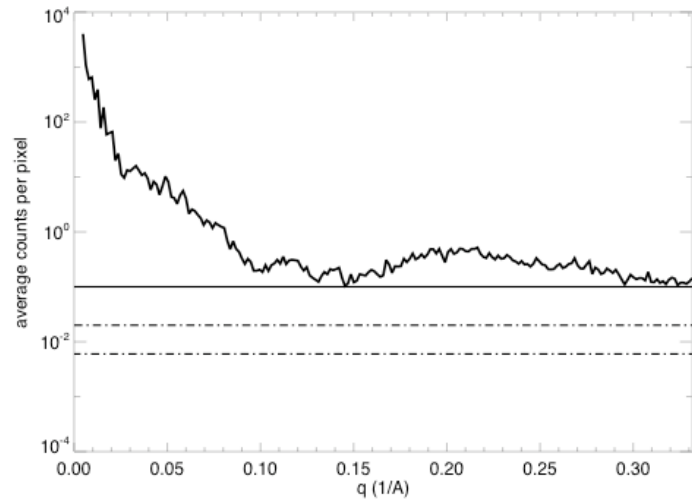
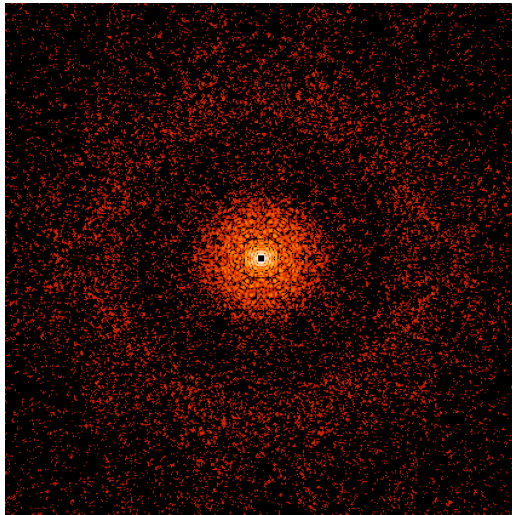


Figure 17. Simulated diffraction data from the CPMV object, for a beam waist diameter of 0.1 μm , while all other parameters kept the same as for Figure 16. The maximum photon count is 5540, excluding the beamstop. The total integrated photons is 3.0×10^5 . The plot shows the radial average of the photon counts. The solid line is at 0.1 counts, the minimum counts needed to classify, and the dotted lines denote the noise level for 10 and 100 averages, for a detector noise of 0.06 photons per pixel.

Summary of detector requirements:

Number of pixels: 1024×1024 , and up to $2k \times 2k$

Noise per pixel: <0.06 photons

Maximum signal: $<10^4$ photons in a pixel (locally $<10^3$ intensity range)

Total data rate: $<10^6$ photons

MTF: $>70\%$ for modulations of period $4p$, where p is the pixel pitch

Hole in detector: 4×4 pixels

Many of these requirements are in line with general directions currently driving the development of X-ray area detectors for synchrotrons and for other experiments planned for the X-ray lasers. The detectors needed for our studies are relatively small, and that alleviates some of the difficulties in developing faster detectors for our studies. The team represented by this LOI will not be developing the diffraction detector, but we will collaborate with laboratories and companies who will do so.

3. LISTS OF INSTRUMENTS

3.1. BEAM DELIVERY, CONDITIONING, AND FOCUSING

- Apodising apertures
- Gas attenuator
- Attenuator foils?
- Pulse compressor/monochromator
- Beam profile monitor
- Rapid beam shutter

OPTICS:

Low-scatter focussing:

- Kirkpatrick-Beaz mirror pair (1 micrometer)
- Kirkpatrick-Beaz mirror pair for fine focus (0.1 micrometer) inside the sample chamber

Require 0.5 to 1 m focal length optic, 0.1 micron focus (<0.1 microrad slope errors)
Entire PSD needs to be controlled (e.g. high-frequency smoothing by ion-beam -- Spiller 2003)
Require high-resolution wavefront sensor
Zone plate

3.2. SAMPLE ENVIRONMENT

- Vacuum Vessel, pressure less than 10^{-6} mbar, ideally 10^{-8} mbar

3.3. RAPID CHANGE ELECTRON CRYO-MICROSCOPY STAGE for samples embedded in vitreous ice or mounted on a support (this will also be used for arrays of particles and crystals on a substrate)

- A three-axis goniometer with cryogenic sample holder will be used for mounting samples. A horizontal drive provides motion along the spindle axis. The main drive provides motion along the Y- and Z-axis and rotation around the X-axis. The samples will be kept at liquid nitrogen temperatures and will be changed without breaking the vacuum system.
- An optical microscope with a long focal length will be used to align the samples and position the rotation axis to the beam.

3.4. PARTICLE INJECTOR SYSTEM

- Sample purification and characterization
- Electrospray ionization (ESI) and quantification of aerial density
- Droplets and electrospray drops (currently ~ 0.1 micron)
- Aerodynamic lenses

3.5. PARTICLE MANIPULATION

- Particle trajectory control and diagnostics
- Ion traps and lasers for capturing and orienting particles in the gas phase
- Particle alignment: >100 μJ , 100 ns pulsed (or 1 kW continuous) near-IR laser.

3.6. DIAGNOSTICS AT THE INTERACTION POINT

- Fluorescence flash monitors to position hits along the X-ray beam path (these may need to provide some spectral resolution)
- Mass spectrometer to analyse fragmentation
- Electron spectrometer
- Inspection microscope and ccd camera

3.7. A PAIR OF X-RAY AREA DETECTORS

Tandem pair, 20-500 mm adjustable distance (independently)

- Channel plate collimator

3.8. DIAGNOSTICS AT THE BEAM DUMP (behind the area detectors)

- X-ray spectrometer to measure pulse shape and intensity distribution on a shot-to-shot basis as a function of wavelength. Could also be used to get total integrate energy.
- High-resolution wavefront sensor
- Zone plate image of the beam (optional)

4. COMPUTER HARDWARE AND SOFTWARE FOR DATA ACQUISITION AND DATA PROCESSING

4.1 Collecting, handling, and storing very large data sets

As in crystallography, the computer is a major instrument component and algorithms are required to generate images from the measurements. Depending on the complexity of the molecule to be studied, the size of a diffraction pattern will be $N \times N$ pixels, with $N = 100$ to 1000 (some applications may require $N = 10,000$). The need to build up a large enough signal to atomic resolution and to obtain full 3D information may require about $10^5 - 10^6$ diffraction patterns to be collected from a series of identical particles. This corresponds to about 40 Tb of data for $N = 1000$. This volume of data could be collected in a day's operation and computing and data handling resources must be able to keep pace with this generation rate.

4.2 Data processing

The processing of the non-crystalline diffraction data, with reconstructed electron density as the end result, consists of two distinct stages. In the first, a large number of noisy 2D diffraction patterns must be given orientations in 3D reciprocal space and merged into a single 3D diffraction pattern. In the second stage, phases are derived from the 3D diffraction pattern based on iterative phase-retrieval methods from the over-sampled diffraction data set.

Image classification: The patterns will be collected from particles at random and unknown orientation, so these patterns must be brought in relation with each other and assembled into a 3D data set. Due to the large number of images required for an adequate signal-to-noise ratio in coherent x-ray setting, it may also be necessary to significantly advance the power of basic algorithms used, e.g. in EM studies. The standard approach has been to divide the merging stage into two operations: classification and orientation. Classification corresponds to applying criterion to establish which 2D diffraction patterns have similar orientations and can therefore be averaged to improve the signal-to-noise ratio.

Image classification can be done successively, starting at low resolution and performing finer groupings as the resolution increases. The requisite algorithms are easily parallelised. Once the diffraction data is classified and averaged to achieve the required signal to noise ratio, it must be assembled into a N^3 array (now about 16 Gb).

Data reduction and merging: The relative orientation of the averaged images may be determined through the method of common lines, a technique widely used in electron microscopy, where the micrographs represent planar sections through the centre of the molecular transform. Diffraction images are different in that they represent spherical sections of centric objects. Each pair of images will intersect in an arc that also passes through the origin of the molecular transform. If the signal is strong enough for the line of intersection to be found in two diffraction images, it will then be possible to establish the relative orientation of these images. We note that due to the curvature of the sections, the common arc will provide a three-dimensional fix rather than a hinge-axis. Moreover, the centric symmetry of the modulus of the molecular transform ensures that we obtain 2 x 2 independent repeats of the common lines in the two images. This feature provides redundancy for determining sample orientation, and is unique to the diffraction geometry. It can lead to a complete determination of the relative angles of the diffraction patterns.

The standard approach to merging, i.e. classification followed by orientation, suffers from some arbitrariness that can be circumvented without compromising performance. Primary among these is the need to establish at the outset the number of orientations that will be averaged and also to choose criteria to define membership in the orientation classes (e.g. closeness to representatives). In an alternative approach, each 2D diffraction pattern will automatically be assigned a unique set of Euler angles determined by a pseudo energy functional defined on an associated "adjacency graph," which is a natural encoding of the likelihood that two patterns are close in orientation given the

applicable model of noise (e.g. Poisson counting statistics). Based on preliminary studies by Elser, it appears that computationally the most challenging step in this approach is not the minimization of the pseudo energy functional but rather the initialization of the minimization, which corresponds to embedding the adjacency graph into the space of orientations so that the topology is correct. The solution to this problem, and in effect the rate-limiting step of the entire approach, is to obtain the four lowest eigenvectors of the adjacency matrix using the Lanczos sparse matrix algorithm. A particularly attractive feature of this initial embedding step is that it simultaneously serves as a diagnostic for data acquisition. From the spectrum of the lowest eigenvalues one can readily assess the degree to which the collected data forms a quasi-continuum in the space of orientations, and assembly of the 3D diffraction pattern can proceed. This diagnostic is fast because it acts on a relatively raw form of the collected data (adjacency graph) and could perhaps be used in "real time" during data collection.

Phasing: The 3D structure is related to the measured amplitudes through a 3D Fourier transform. Since only the Fourier magnitudes are measured, phase retrieval methods must be employed to derive the phases and so determine the structure. The most promising approach is a method of generalised projections, in which known constraints are iteratively applied. This requires many thousands or tens of thousands of iterations, with two or more N^3 FFTs per iteration. The FFT is a demanding algorithm for parallel machines, and performance is usually limited by communication speeds. Efficient routines can be achieved when the number of nodes is equal to N , in which case two across-processor transposes are required. In general, the reconstruction problem requires 1024^3 FFTs to be performed in under a second.

Hardware: A dedicated computer hardware configuration will be needed with large storage capacity. The primary objective will be to orient the patterns. Because the data rate in the experiments is very high (around 40 Tb/day), this will be a computing cluster where each node is responsible for either identifying members of particular orientation classes or establishing connectivity in the adjacency graph. A separate cluster will be used in the later steps of structure determination.

5. TECHNOLOGY DEVELOPMENT IS REQUIRED NOW IN THE FOLLOWING AREAS:

- Sample preparation: including membrane nanocrystals, arrays of particles, virus templates, pure protein samples, viruses, cells,
- Electrospray delivery, including trapping and storing
- Droplet delivery, including micro-scale nozzles
- Laser orientation of particles
- Debris control, diagnostics of XFEL-particle interaction
- X-ray pulse compression; pump-probe optics
- Algorithms for classification and orientation of terabyte datasets; reconstruction
- Experiment and MD simulation of structural variability of proteins in vacuum and injection
- Prediction and validation of XFEL-molecule interaction

6. ESTIMATED COSTS

Sample Handling	FTE	Cost
Cryo-EM style sample prep (vitreous ice)	3	175 k€
Goniometer	0.5	150 k€
Dip-pen sample arrays; labeled viruses; nanodots	4	150 k€
Pump-probe sample arrays	1	50 k€
Chamber, motors & diagnostics, control	6	400 k€

Monochromator	2	150 k€
Sample Handling - Spraying		
1st Sample selection and injection system:		
Electrospray injector; ion mobility; ion trap	8	550 k€
Aerodynamic lens	2	50 k€
Droplet & tamper	4	400 k€
Test and development	3	50 k€
Mass-spec, diagnostics development	2	200 k€
Chamber and diagnostics	3	200 k€
Alignment laser (100 ns, 100 mJ)		200 k€
Alignment test and development	7	250 k€
2nd Sample selection and injection system (as above, for testing conditions outside the hutch in a test laboratory):		1850 k€
Optics		
K-B mirrors, mount	3	1000 k€
Wavefront sensor	1.5	150 k€
High-frequency smoothing	2	75 k€
Characterization	2	50 k€
Spectrometer	3	200 k€
Apodising apertures	0.2	50 k€
Collimating channels	2	300 k€
Reconstruction, Modeling, Software		
XFEL-matter interaction	3	10 k€
Reconstruction tests	3	10 k€
Database	3	50 k€
Magnetic traps and electronics:		200 k€
Optical tweezers:		100 k€
Sample environment for single particles:		200 k€
Detectors:		
Two detector sets (to be decided):	3,000 - 5,000 k€	
Additional electronics & detectors (e.g. fluorescence):		400 k€
Femtosecond laser (microJ/pulse):		380 k€
INSTRUMENTS FOR SAMPLE PREPARATION LABORATORY:		
Centrifuge		50 k€
Ultracentrifuge		150 k€
Chromatography systems		150 k€
Spectrophotometers, electrophoresis, etc.		50 k€
Equipment for cell growth and cell preparation		50 k€
JEOL JEM-2010F HT electron microscope or equivalent		770 k€
JEOL cryo stage and transfer system		170 k€

Microscopes	60 k€
Freezers/Fridges	10 k€
Cold room	100 k€

Acknowledgements

This work was supported by The Swedish Foundation for Strategic Research and under the auspices of the U.S. Department of Energy (DOE) under Contract W-7405-Eng-48 to the University of California, Lawrence Livermore National Laboratory (the project 05-SI-003 was funded by the Laboratory Directed Research and Development Program at LLNL).

7. REFERENCES

- Andersson, I., Large structures at high resolution: spinach ribulose-1,5-bisphosphate carboxylase/oxygenase at 1.6 Å resolution. *J. Mol. Biol.* 259, 160-174 (1996).
- Bates, R. H. T. Fourier phase problems are uniquely solvable in more than one dimension. 1. Underlying theory, *Optik* 61, 247-262 (1982).
- Bergh, M., Timneanu, N., van der Spoel, D.: A Model for the Dynamics of a Water Cluster in an X-ray Free Electron Laser Beam. *Phys. Rev. E* 70 Art. No. 051904 (2004)
- Chapman, H.N., Barty, A., Marchesini, S., Hau-Riege, S.P., Cui, C., Howells, M.R., Rosen, R., He, H., Spence, J.C.H., Weierstall, U., Beetz, T., Jacobsen, C., Shapiro, D. High-resolution ab initio Three-dimensional X-ray Diffraction Microscopy. *J. Opt. Soc. Am. A*, (2005).
- Chapman, H.N., Barty, A., Bogan, M.J., Boutet, S., Frank, M., Hau-Riege, S.P., Marchesini, S., Woods, B., Bajt, S., London, R.A., Plönjes-Palm, E., Kuhlmann, M., Treusch, R., Düsterer, S., Tschentscher, T., Schneider, J.R., Spiller, E., Möller, T., Bostedt, C., Hoener, M., Shapiro, D.A., Hodgson, K.O., van der Spoel, D., Burmeister, F., Bergh, M.L.A., Caleman, C., Huidt, G., Seibert, M.M., Maia, F. R. N. C., Lee, R.W., Szöke, A., Timneanu, N., Hajdu, J. Ultrafast coherent diffraction imaging with a soft X-ray free-electron laser. Submitted (2006).
- Elser, V. <http://arxiv.org/pdf/physics/0505174> Reconstruction of an object from its symmetry-averaged diffraction pattern (2005).
- Fienup, J. R. Phase retrieval algorithms-a comparison, *Appl. Opt.* 21, 2758-2769 (1982).
- Frank, J., Three-Dimensional Electron Microscopy of Macromolecular Assemblies. Academic Press, San Diego (1996).
- Hajdu, J., Hodgson, K., Miao, J., van der Spoel, D., Neutze, R., Robinson, C.V., Faigel, G., Jacobsen, C., Kirz, J., Sayre, D., Weckert, E., Materlik, G., Szöke, A. Structural studies on single particles and biomolecules. In *LCLS: The First Experiments*. pp. 35-62. Published in 2000 by SSRL, SLAC, Stanford, USA (2000).
- Hajdu, J. & Weckert, E. Life Sciences - Scientific applications of XFEL radiation. In *TESLA, the Superconducting Electron-Positron Linear Collider with an integrated X-ray Laser Laboratory*. Technical Design Report., DESY, ISBN 3-935702-00-0, Volume 5, pp. 150-168 (2001).
- Hajdu J. Imaging the structure of single particles and biomolecules. Chapter 2.2.3, *TESLA XFEL, First stage of the X-Ray Laser Laboratory*. DESY, ISBN 3-935702-10-8, pp16-20 (2002).
- Hau-Riege, S.P., London R.A., Szöke, A. Dynamics of X-Ray Irradiated Biological Molecules, *Phys. Rev. E* 69, 051906 (2004a).
- Hau-Riege S. P., Szoke, H., Chapman, H.N., Szoke, A., Marchesini, S., Noy, A., He, H., Howells, M.R., Weierstall, U., Spence, J.C.H., SPEDEN: Reconstructing single particles from their diffraction patterns. *Acta Cryst. A* 60, 294-305 (2004b).
- Hau-Riege S.P., London R.A., Huidt G., Chapman H.N. Pulse requirements for x-ray diffraction imaging of single biological molecules. *Phys. Rev. E* 71, Art. No. 061919 (2005).
- Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quart. Rev. Biophys.* 28, 171-193 (1995).
- Henderson, R. Cryoprotection of protein crystals against radiation-damage in electron and X-ray diffraction. *Proc. R. Soc.* 241, 6-8 (1990).
- Huidt, G., Szöke, A., Hajdu, J., Diffraction imaging of single particles and biomolecules. *J. Struct. Biol.* 144, 219-227, (2003).
- Jayasinghe, S.H., Eagles, P.A.M., Qureshi, A.N. Electric field driven jetting: an emerging approach for processing living cells. *Biotechnol. J.* 1, 86-94 (2006).
- Jurek, Z. G. Faigel, and M. Tegze, Dynamics in a cluster under the influence of intense femtosecond hard x-ray pulses, *Euro. Phys. J. D* 29, 217-229 (2004a).
- Jurek, Z., G. Oszlanyi and G. Faigel, Imaging atom-clusters by hard x-ray free electron lasers, *Europhys Lett.*, 65, 491-497, (2004b).

- Larsen, J. J., Hald, K., Bjerre, N., Stapelfeldt, H., Seidman, T., Three Dimensional Alignment of Molecules Using Elliptically Polarized Laser Fields, *Phys. Rev. Lett.* **85**, 2470 (2000).
- Lee, R. W., Cauble, R. C., Landen, O. L., Wark, J. S., Ng, A., Rose, S. J., Lewis, C., Riley, D., Gauthier, J.-C., Audebert, P. Plasma and Warm Dense Matter Studies. LCLS: The First Experiments. pp. 35-62. Published by SSRL, SLAC, Stanford, USA. 2000.
- Lee RW: Plasma physics experiments. Chapter 2.2.2., In the TESLA XFEL, First stage of the X-Ray Laser Laboratory. DESY, ISBN 3-935702-10-8, pp. 11-16 (2002).
- Marchesini, S., Chapman, H. N., Hau-Riege, S. P., London, R. A., Szöke, A., Coherent X-ray diffractive imaging: applications and limitations, *Opt. Express* **11** (19), 2344-2353, (2003a).
- Marchesini, S., He, H., Chapman, H. N., Hau-Riege, S. P., Noy, A., Howells, M. R., Weierstall, U., Spence, J. C. H., X-ray image reconstruction from a diffraction pattern alone, *Phys. Rev. B* **68** (114), 140101, (2003b).
- Miao, J., Sayre, D., Chapman, H.N., Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *J. Opt. Soc. Am.* **15** 1662-1669 (1998).
- Miao, J., Charalambous, P., Kirz, J. & Sayre, D, Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens, *Nature* **400**, 342-344 (1999).
- Miao, J., Hodgson, K., Sayre, D., An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images. *Proc. Natl. Acad. Sci. USA* **98** (12), 6641-6645, (2001).
- Miao, J., Ishikawa, T., Johnson, B., Anderson, E. H., Lai, B., Hodgson, K. O., High resolution 3D X-ray diffraction microscopy, *Phys. Rev.* **89**, 088303, (2002).
- Miao, J., Hodgson, K. O., Ishikawa, T., Larabell, C. A., LeGros, M. A., Nishino, Y., Imaging whole Escherichia coli bacteria by using single particle x-ray diffraction, *Proc. Natl. Acad. Sci. USA*, **100** (1), 110-112, (2003).
- Miao, J., H.N. Chapman, J. Kirz, D. Sayre and K.O. Hodgson, *Annu. Rev. Biophys. Biomol. Struct.* **33**, 157-176 (2004).
- Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, Potential for biomolecular imaging with femtosecond X-ray pulses, *Nature*, **406**, pp. 752-757 (2000).
- O'Konski, C. T. and Zimm, B. H., *Science* **111**, 113 (1950)
- Persson, P., Lunell, S., Szöke, A., Ziaja, B. & Hajdu, J. Shake-up and shake-off excitations with associated electron losses in X-ray studies of proteins. *Protein Science* **10**, 2480-2484 (2001).
- Reiss, H. R. Physical basis for strong-field stabilization of atoms against ionization. *Laser Physics* **7**, 543-550 (1997).
- Robinson, I. K., Vartanyants, I. A., Williams, G. J., Pfeifer, M. A., Pitney, J. A., Reconstruction of the shapes of gold nanocrystals using coherent x-ray diffraction, *Phys. Rev. Lett.*, **87** (19), 195505, (2001).
- Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., Dobson, C. M., Robinson, C. V. Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc. Natl. Acad. Sci. USA* **97**, 5185-5190 (2000).
- Ruotolo, B.T., Giles, K., Campuzano, I., Sandercock, A.M., Bateman, R.H., Robinson, C.V. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **310**, 1658-1661 (2005).
- Saalmann U, Rost JM: Ionization of clusters in strong x-ray laser pulses. *PHYS REV LETT* **89** (14): Art. No. 143401 SEP 30 (2002)
- Santra R. and Greene, C.H.: Xenon Clusters in Intense VUV Laser Fields. *Phys. Rev. Lett.* **91**, 233401 (2003)
- Sayre, D. Some implications of a theorem due to Shannon. *Acta Cryst.* **5**, 843 (1952).
- Sayre, D. in *Proceedings of the NATO Course, Erice* (H. Schenk, ed.) pp.353-356 (1990).
- Sayre, D., X-ray crystallography. *Struct. Chem.* **13** (1), 81-96 (2002).
- Sayre, D., Chapman, H.N., X-ray microscopy. *Acta Cryst. A.* **51**, 237-252 (1995).

- Sayre, D., Chapman, H. N. & Miao, J. On the extendibility of x-ray crystallography to noncrystals. *Acta Cryst. A*, 54, 232-239 (1998).
- Seideman, T. On the dynamics of rotationally broad, spatially aligned wave packets. *J. Chem. Phys.* **115**, 5965-5973 (2001)
- Siegman, A. E. "Lasers," University Science Books (1986).
- Starodub, D., Doak, R. B., Schmidt, K., Weierstall, U., Wu, J. S., Spence, J. C. H., Howells, M., Marcus, M., Shaprio, D., Barty, A., and Chapman, H. N., Damped and thermal motion of laser-aligned hydrated macromolecule beams for diffraction, *J. Chem. Phys.* **123**, 244304 (2005).
- Szöke, A. X-ray and electron holography using a local reference beam, in *Short Wavelength Coherent Radiation: Generation and Application*. Attwood, D. T. and Bokor, J. eds., American Institute of Physics Conference Proceedings No. 147, New York (1986).
- Szöke, A. Holographic methods in X-ray crystallography. 2. Detailed theory and connection to other methods of crystallography. *Act. Cryst.* A49, 853-866 (1993).
- Szöke, A., Szöke, H., Somoza, J.R. Holographic methods in X-ray crystallography 5. Multiple isomorphous replacement, multiple anomalous dispersion and non- crystallographic symmetry. *Acta Cryst.* A53, 291- 313 (1997).
- Szöke, A. Time-resolved holographic diffraction at atomic resolution. *Chem. Phys. Letts.* 313, 777-788 (1999).
- Tegze, M., Faigel, G. Atomic-resolution X-ray holography. *Europhys. Lett.*, 16, 41-46 (1991).
- Tegze, M., Faigel, G. X-ray holography with atomic resolution. *Nature* 380 49-51 (1996).
- Timneanu, N., Coleman, C., Hajdu, J., van der Spoel, D.: Auger electron cascades in water and ice. *Chemical Physics* 299, 277-283 (2004)
- Tito, M. A., Tars, K., Valegard, K., Hajdu, J., Robinson, C V. Electrospray Time-of-Flight Mass Spectrometry of the Intact MS2 Virus Capsid, *J. Am. Chem. Soc.* 122, 3550-3551 (2000).
- van Heel, M., Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., Patwardhan, A., Single-particle electron cryo-microscopy: towards atomic resolution. *Quart. Rev. Biophys.* 33, 307-269 (2000).
- Wabnitz H, Bittner L, de Castro ARB, et al. Multiple ionization of atom clusters by intense soft X-rays from a free-electron laser *NATURE* 420 (6915): 482-485 DEC 5 2002
- Wang, Q. *et al.*, Icosahedral virus particles as addressable nanoscale building blocks. *Angew. Chem. Int. Ed.* **41**, 459-462 (2002).
- Ziaja B, van der Spoel D, Szoke A, Hajdu, J: Auger-electron cascades in diamond and amorphous carbon *PHYS REV B* 64 (21): Art. No. 214104 (2001).
- Ziaja B, Szoke A, van der Spoel D., Hajdu, J: Space-time evolution of electron cascades in diamond *PHYS REV B* 66 (2): Art. No. 024116 (2002).
- Ziaja, B. London, R.A., Hajdu, J.: Unified model of secondary electron cascades in diamond. *J. Appl. Phys.* 97, 064905 (2005).
- Ziaja, B. London, R.A., Hajdu, J.: Ionization by impact electrons in solids: Electron mean free path fitted over a wide energy range. *J. Appl. Phys.* 99, 033514 (2006).

